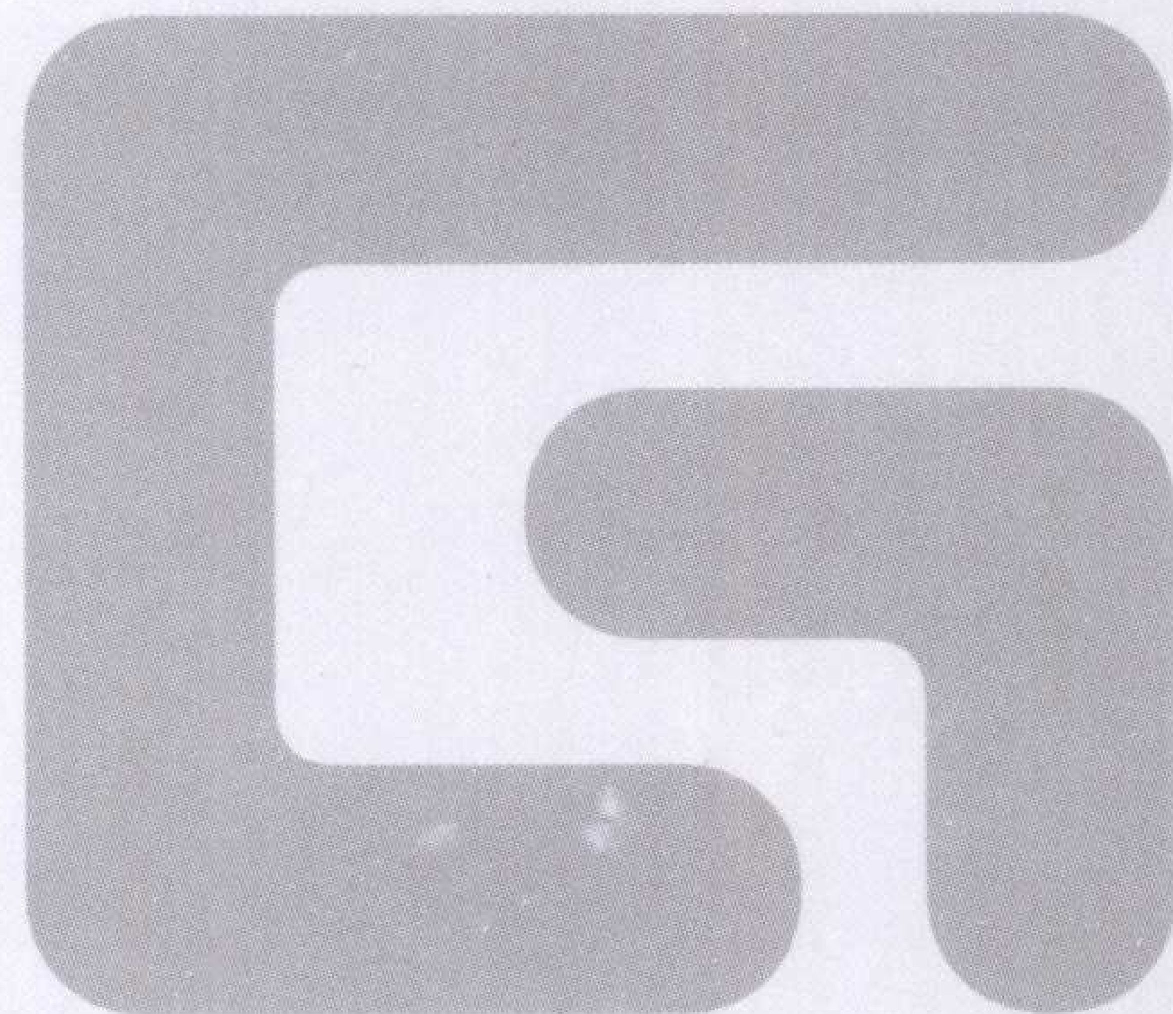
万卷方法 | 社会科学研究方法经典译丛

CRC Press
Taylor & Francis Group

GENERALIZED LATENT VARIABLE MODELING: MULTILEVEL, LONGITUDINAL, AND STRUCTURAL EQUATION MODELS



# 广义潜变量模型

## ——多层次、纵贯性以及结构方程模型

Anders Skrondal
Sophia Rabe-Hesketh 著

陈华珊 叶鹏飞 李 洋 译



重庆大学出版社
http://www.cqup.com.cn

GUANGYI QIANBIANLIANG MOXING:
DUOCENGCI ZONGGUANXING YIJI JIEGOU FANGCHENG MOXING

该书综合以及扩展了各种潜变量模型，包括多层次模型、广义线性混合模型、纵贯性模型、题器响应模型、潜类模型以及结构方程模型。通过对潜变量模型的介绍，作者清晰地解释及比较了生物统计学、心理计量学、经济计量学以及统计学对模型估计及预测的方法。该书同时向研究人员展现了如何应用潜变量模型在社会学、经济学、心理学、医学等不同学科去解决实质性问题，该书所提供的经验案例详细具体，并包含有软件计算过程及数据，非常适合学习。



上架建议：学术社科

ISBN 978-7-5624-5393-2

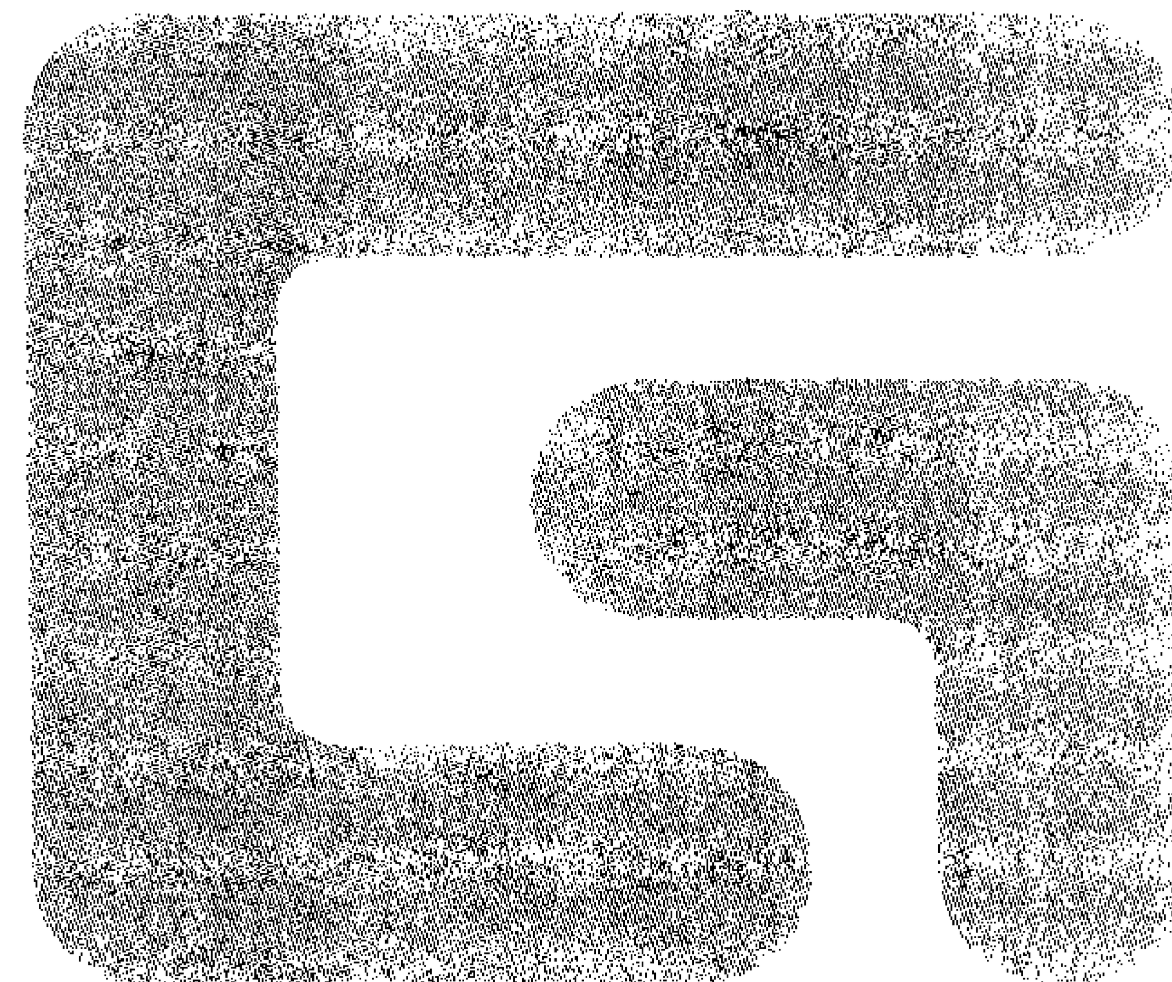


定价: 69.00元

万卷方法 | 社会科学研究方法经典译丛

■ 主编 沈崇麟 夏传玲

GENERALIZED LATENT
VARIABLE MODELING:
MULTILEVEL, LONGITUDINAL,
AND STRUCTURAL
EQUATION MODELS



# 广义潜变量模型

## ——多层次、纵贯性以及结构方程模型

Anders Skrondal
Sophia Rabe-Hesketh 著

陈华珊 叶鹏飞 李 洋 译

重庆大学出版社

[redacted]

**图书在版编目(CIP)数据**

广义潜变量模型:多层次、纵贯性以及结构方程模型/(美)斯科隆多(Skrondal, A.),(美)赫斯基思(Hesketh, S. R.)著;陈华珊,叶鹏飞,李洋译.—重庆:重庆大学出版社,2011.1

(万卷方法)

书名原文:Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models

ISBN 978-7-5624-5393-2

Ⅰ.①广… Ⅱ.①斯…②赫…③陈…④叶…⑤李… Ⅲ.①统计模型—研究 Ⅳ.①C81

中国版本图书馆CIP数据核字(2010)第089436号

**广义潜变量模型**

——多层次、纵贯性以及结构方程模型

(美)安德斯·斯科隆多(Anders Skrondal)
(美)索菲亚·拉贝-赫斯基思(Sophia Rabe-Hesketh) 著
陈华珊 叶鹏飞 李 洋 译
策划编辑 雷少波 林佳木
责任编辑:李定群 高鸿宽 版式设计:林佳木
责任校对:贾 梅 责任印制:赵 晟

*

重庆大学出版社出版发行
出版人:邓晓益
社址:重庆市沙坪坝正街174号重庆大学(A区)内
邮编:400030
电话:(023) 65102378 65105781
传真:(023) 65103686 65105565
网址:http://www.cqup.com.cn
邮箱:fxk@cqup.com.cn (营销中心)
全国新华书店经销
自贡新华印刷厂印刷

*

开本:787×1092 1/16 印张:24.25 字数:619千 插页:16开2页
2011年1月第1版 2011年1月第1次印刷
印数:1—4 000
ISBN 978-7-5624-5393-2 定价:69.00元

# 作译者简介

安德斯·斯科隆多(Anders Skrondal)是英国伦敦经济学院统计学系教授。

索菲亚·拉贝-赫斯基思(Sophia Rabe-Hesketh)是美国加州伯克利大学教育学院教授及生物统计学系教授。

陈华珊:中国社会科学院社会学所助理研究员,从事社会学方法论及组织社会学研究。

叶鹏飞:中国劳动关系学院工会学系讲师,清华大学博士毕业,主要研究领域是劳工和城市社会学。

李洋:北京市社会科学院社会学研究所助理研究员,清华大学博士毕业。

陈华珊翻译第 5 章至第 11 章以及第 13 章和第 14 章。叶鹏飞翻译了第 1 章、第 2 章和第 3 章。李洋翻译了第 4 章和第 12 章。

## 总序

# 社会研究方法的现状及其发展趋势

近年来,社会调查技术和社会研究方法都有很大的发展。在调查技术方面,自20世纪70年代以来,社会变迁多次横断面的跟踪调查研究,几乎成为所有国家和地区了解社会结构转变和社会发展状况的基础性调查。这种调查不仅对社会学的研究有很大促进作用,而且对整个社会科学的研究都产生了重大影响,并且这些调查结果有的已作为政府有关部门决策的重要依据。国际上比较著名的此类调查有:美国芝加哥大学全国民意调查中心(National Opinion Research Center,简称NORC)的“社会综合调查(General Social Survey ,简称GSS)”,英国埃塞克斯大学调查中心进行的“全国家庭生活和社会变迁调查”,法国经济和社会调查所进行的“全国经济社会调查”,日本社会学会组织进行的“全国社会分层与社会流动调查(简称SSM)”。中国台湾“中央”研究院社会学研究所,也每两年进行一次“台湾社会变迁基本调查”。美国的“社会基础调查”,现在已成为年度性的调查项目,它是美国国家基金会目前资助的最大的社会科学研究项目。以上这些调查,除美国的调查外,一般均因经费原因采用纵向的间隔性重复调查法,即每隔一段时间,进行一次全国规模的抽样调查。每次调查除保留社会研究所需的基本项目外,都有不同的主题。在间隔若干时间后,再重复同一主题的调查,这样的研究设计,使社会变迁研究在可以涉及更为广泛的研究领域的同时,具有更好的积累性和可比性。多年来,这些基础性调查获得的资料,滋养着大批的社会科学研究者,有时一项调查就有很多名博士生用来写博士论文,以此取得的研究成就,其可靠性受到社会科学界的广泛认同。例如1997年出版,以台湾地区社会变迁基本调查数据为基础的研究报告集《90年代的台湾社会,社会变迁基本调查研究系列二》收集论文16篇,内容涉及社会生活的各个方面,在台湾地区引起了极大的反响。

国内社会科学界在这方面也有了长足的发展。笔者所在的中国社会科学院社会学研究所的社会调查和方法研究室,组织或参与了多项与社会变迁有关的大规模抽样调查,取得了一定的研究成果,并积累了大量有关社会变迁的宝贵数据资料,其中主要有:

1. 城乡家庭变迁系列调查:该课题是由中国社会科学院社会学研究所牵头,联合北京大学和地方社科院的研究人员展开的一项类似多次横断面的城乡家庭变迁调查。这一调查始于1981年的“中国五城市婚姻家庭调查”,而后有1988年的“中国农村家庭调

查”、1991 年的“中国七城市家庭调查”、1998 年的“中国城乡家庭变迁调查”。

2. 有关中国城乡社会变迁的系列调查：这一调查始于 1991 年的第二批国情调查，然后有 1992 年的“中国城乡居民生活调查”、1993 年的“第三批国情调查”、1995 年的“第四批国情调查”和 1997 年的“中国沿海发达地区社会变迁调查”。上述调查虽然还不是严格意义上的多次横断面的纵贯研究，但研究者已在研究设计中尽量考虑到纵贯研究的基本原则，如调查队伍的稳定、指标的可比性和样本空间的延续性等。

3. 中国城乡社会变迁调查：这一调查始于 2000 年，为中国社会科学院重大课题。目前已经完成第一期第一次调查和第二次调查，今后将把这一调查发展为连续的、定期进行的社会变迁调查。

在纵向调查技术取得长足进步的同时，20 世纪末至今，电话调查也有很大发展。电话调查涉及的范围几乎与个别（面对面）访谈同样全面。电话调查中使用的一系列方法，是在 20 世纪 70 年代后期和面对面调查一起发展起来的。在 20 世纪 80 年代中期，电话调查开始变得很普遍，并且成为许多场合中各种调查方法的首选。正如某些学者所言，一种在公共和私营部门被人们用来帮助提高决策效率的收集信息的有效方法为人们所普遍认同时，这一现象本身就具有方法论上的意义。不仅如此，电话调查还有很大的实践意义，因为它为研究者提供了更多的控制调查质量的机会。这一机会包括抽样、被调查人的选择、问卷题项的提问、计算机辅助电话访谈（CATI）和数据录入。正因为如此，今天在各种社会调查中，如果没有发现其他重要的足以放弃使用电话调查的原因，电话调查由于其独特的对调查质量进行全面监控的优点，常常成为各种调查方式的首选。由笔者翻译，重庆大学出版社出版的《电话调查方法：抽样、选择和督导》一书，也于 2005 年面世。

无论是纵向调查抑或电话调查，实际上都是收集研究资料的方法，而应用社会科学的发展，不仅在于调查技术，即收集资料技术的发展，还在于研究方法和分析技术的发展。近年来，无论是定性研究方法，还是定量研究方法都有了长足的发展。

首先，计算机技术的发展可谓突飞猛进，它对当今社会生活的各个方面产生了巨大的影响，在悄悄地改变着社会科学的研究风格和研究方式的同时，也大大提升了社会科学学者的研究能力。这种影响表现在研究过程的各个阶段，从理论建构（概念映射）、问卷设计（专业的问卷设计软件）、调查实施（计算机辅助访谈、计算机辅助电话访问系统、网络在线调查系统）、数据录入（光学标记识别软件）到数据分析（包括文本、声音、图像资料的处理），甚至延伸到写作发表阶段。这样的过程发生在如社会学、经济学、政治学、心理学、教育学中，促进了学科之间的相互借鉴和交叉融合，至少在研究方法上呈现出这种趋势。随着计算机计算能力的大幅度提高，20 世纪 80 年代后期，统计学领域内发生了一场“革命”，主要表现在对定类和定序变量的建模能力的大幅度提高上，以及与分布无关的统计分析模型的发展之上，特别是基于“Resampling”（包括 Bootstrap、Jackknife、Monte Carlo 模拟等）的建模技术。同时，计算能力的提高还带动了基于神经网络、动态模拟、人工智能、生态进化等新兴的分析和预测模型的发展。这些进展都为定量社会科学研究提供了更多的可供选择的工具。

亚德瑞安·E. 拉夫特里（Adrian E. Raftery）依据社会学家所处理的数据类型，将定量社会学在美国的发展划分为三个时代：第一代起始于 20 世纪 40 年代，交互表是其主要处理对象，研究重点是关联度和对数线性模型；第二代起始于 20 世纪 60 年代，主要处理单层次的调查数据，Lisrel 类型的因果模型和事件史分析是其研究重点；第三代起始于 20 世纪 80 年代后期，开始处理诸如文本、空间、社会网络等非传统的数据类型，目前尚没有

形成成熟的形态。拉夫特里的综述,虽然更强调定量社会学研究对统计学的贡献,但也大致勾勒出定量社会学在国外的发展脉络。

从分析模型的角度来看,定量分析在以下几个方向有了突破性发展:

1. 缺失值处理:由于社会生活的复杂性,社会调查数据常常出现缺失值,传统的处理方式是忽略这些缺失值,或者用均值替代。但现在则倾向于用多重插值法(multiple implation)或者其他基于模型的方法进行处理。这些技术的发展,不仅会增强我们对数据的处理能力,而且将改变我们设计问卷的方式。基于这些技术,我们在不增加被访者负担的前提下,大大增加了调查问卷的内容:每个被访者只回答问卷的一部分,然后通过对缺失值的处理,获得他们对未回答部分的估值。

2. 非线性关系:线性假定是经典定量分析的一个常见假定,但在实际研究当中,线性假定只能被看作是对社会现实的一个逼近和简化。面对具体的研究数据,如果没有理论上的明确指引(不幸的是,我们常常没有中程理论的指引),我们是无法在线性模型和非线性模型之间作出取舍的。但 MARS 模型的出现,让我们可以从经验数据当中获得最为拟合的变量之间的函数关系,而不必预先作出线性假定。这样,理论思考和数据分析就可以实现一个互动的循环过程,定量分析就不单单是对理论和假设的简单证伪过程,而是理论思维一个重要组成部分。

3. 测量层次:20 世纪六七十年代的统计模型,大多要求数据的测量层次在定距以上,如因素分析,但社会学的调查数据却大多为定类或定序数据。对应分析、Loglinear、Logit、Logistic Regression、潜类分析、Ordinal Regression、Normal Ogive Regression 等统计模型的出现,大大提高了定量社会学处理定类和定序数据的能力。

4. 测量模型:基于文化、社会、心理和认知等方面的考虑,在社会学界仍有人对问卷调查在中国的效度提出质疑。抛弃"本土化"的文化执著,我们更应当关注的是问卷调查的项目反应理论(item response theory),即被访者回答问卷题器时的过程模型。这方面的进展主要表现在两个方面:一是分解测量量表的成分,如 Rasch model、IRT 分析、Mokken 分析等;二是将测量模型与因果模型或其他分析模型结合在一起,明确把测量误差引入到分析当中,充分评估它们对分析结果的影响,如结构方程模型。

5. 潜变量模型:与测量模型相关联的另外一个发展方向是潜变量模型,例如,潜变量分层分析(latent class analysis)、潜变量结构分析(latent structure analysis)、潜变量赋值分析(latent budget analysis)等。"潜变量"这一概念表明,我们可以通过测量"显变量"来测量无法直接观察的理论概念,如权力、声望、地位等。这样,理论和现实之间,通过"潜变量"到"显变量"的映射(测量过程),就有了连接的桥梁。

6. 分析单元的层序性:在定量分析当中,我们常常强调要避免出现"生态谬误",即分析单元的层次和结论或推论的层次不一致。与其相关的方法论争论是"宏观和微观"的问题。随着多层次模型的出现,我们可以同时考察多个层次上的问题,我们可以把个人放在其家庭背景中,再把家庭放在社区的背景下,考察个人层次的变量对社区变量的效应,或者社区层次的变量对个体行为的具体影响。在定量分析模型当中,"宏观和微观"的连接获得了建模技术上的支持。在这个领域当中,还有一个方向也值得关注:分析宏观层次的数据,对微观层次进行推论。

7. 社会网络模型:区分"关系数据"和"属性数据",是把分析重点从个体/群体等社会单元转移到这些社会单元之间关系的第一步,社会网络模型是目前发展较快的一个定量分析领域,其理论根基是结构主义。社会网络分析目前仍然具有较浓厚的"形态学"特征(基于图论的缘故),但却为我们理解社会关系在社会空间上的形态奠定了基础,通过计

算机模拟和研究社会网络的历期数据，研究社会结构的"发生学"性质模型也处在萌芽状态当中。

8. 系统动力学：如果说社会网络模型是在社会空间上拓展定量社会学的研究手段，那么社会过程在时间上和物理空间上的属性，则是事件史模型、事件数模型、历期分析、Cox 回归、时间序列分析、Cohort 分析、状态空间模型等模型的研究对象。在这个领域，计量经济学为定量社会学研究提供了许多有益的范例。

9. 预测模型：上述模型仍然是在分析主义的范式下。有些社会学的应用研究，更强调模型的预测精度，而不是模型的认知价值，例如，社会趋势的预测。由于计算能力的提高，神经网络、基因算法、人工智能、模式识别等数据挖掘技术有了长足发展，已经出现了许多拟合经验数据的预测模型，比较成功的应用出现在计量经济学领域（如对股市的预测）。

10. 计算机模拟：对于社会学应用研究而言，研究的对象具有历史性、规模大、变迁的过程不仅漫长且表现某种渐进性的特点，且因社会隔离/社会伦理原因无法接近或有实验禁忌等，无法直接进行观察和研究，这时计算机模拟就成为一个可供选择的替代方案。计算机模拟主要有两个类型：一是基于计算机网络的模拟：每台微机作为一个代理，整个网络作为"社会"实时演化，如法国的 Swarm 计划；二是基于概念模型的系统，在计算机时间上，按照既定规则运行，较有名的研究是罗马俱乐部的《增长的极限》，常见的软件有 Simul，Arena 等。自然科学家对此方向似乎比社会学家更有兴趣。

定性研究方法一直是社会学研究领域中比较传统的研究方法，在社会学研究的古典时期，它甚至是社会学家手中唯一的研究方法。但随着定量研究方法在社会学研究中的广泛应用，定性研究方法就似乎越来越不受人们的重视。但需要澄清的事实是，在定量分析模型取得飞速发展的同时，在过去的二十多年里，定性研究方法也有了长足的进步。主要表现在以下六个方面：

1. 研究素材日益扩大：除了传统的参与观察、深度访谈、专题小组访谈之外，会话、交谈、电视、广播、文档、日记、叙事、自传（autobiography）等社会过程中自然产生的素材，甚至社会学理论本身（理论的形式化），也开始进入定性分析的视野当中。所有这些资料，不仅可以以文本的格式存储，而且，新型的多媒体介质，如图像、声音和视频，作为原始的分析素材，也日益成为定性分析的新宠。

2. 分析方法更加多样：定性方法的种类在最近的二十多年中，更是有了一个质的飞跃。在比较传统的、源自语言学的方法，如内容分析、话语分析、修辞分析、语意分析、符号学、论据分析等方法之外，社会学家也创造出自己独特的定性分析方法，如施特劳斯（Strauss）等人的扎根理论、海斯（Heise）的事件结构分析、拉津（Ragin）的定性对比分析、Abbott 和 Hrycak 采用最优匹配技术的序列分析、亚贝儿（Abell）的形式叙事分析（formal narrative analysis）、鲍尔（Bauer）等人的语库建设、Attride-Stirling 等人的主题网络分析和神经网络技术应用的定性分析领域。所有这些方法的一个共同特征是，把定性研究向更加系统、更加精确、更加严格、更加形式化的方向推进。

3. 认识论基础更加多元化：现象学、释义学和本土方法论（ethnomethodology）的认识论，一直是定性分析的大本营，但近年来，实证主义也开始逐渐为定性分析所接纳，解释和阐释之间，由激烈的对立关系，逐渐演变为相互融洽的关系。

4. 研究过程更加客观规范：定性分析的一个主要问题在于阐释过程中不可避免的主观性。为了尽可能消除"解释者偏见"和主观选择性，定性分析开始遵循严格的程序模板或程序规则，并尝试引入定量分析中的"信度""效度""代表性"等概念，通过编码和对比，再加上传统的定

性分析标准，如可解释性、透明性和一致性，使得定性研究的过程更加规范、阐释的结果更加客观，研究的结论更加可信。

5. 研究过程更加有效率：这主要应归功于大量计算机辅助定性数据分析(CAQDA)软件的涌现。从20世纪80年代以来，定性分析过程的数字化和计算机化，已经是一个不可逆转的大趋势。这种发展趋势与定性研究者的理论取向无关，不管他们的理论立场是实证主义、符号互动论，还是本土方法论，大多数定性研究者都在自己的研究当中，开始采用计算机来辅助定性资料的分析过程。据不完全统计，目前已经有二十多种定性分析的软件，分别隶属于德国、英国、法国、美国等国家。其中，有一些软件是国外研究机构的科研成果，可以免费使用，但比较成熟的定性辅助系统大多是商业软件。这些定性分析的辅助系统，不仅使得研究者从处理大量文字材料的繁复劳动中解放出来，而且能够让研究者共享他们各自分析的细节，从而改变定性研究的流程和研究集体之间的合作方式。同时，由于采用数据库结构，定性资料的管理也更加方便，这就为组织大型定性研究项目(包括多个研究地点、多个研究对象、历时的定性研究)提供了新的可能性。越来越多的定性研究人员开始走出他们的摇椅，坐到计算机屏幕前、湮没在访谈资料和故纸堆中的定性社会学家的形象已经一去不复返了。

6. 定性研究和定量研究的结合更加紧密：在定量分析方法的教材中，定性研究常常被看做是定量研究的前期准备工作，但定性研究者却持完全相反的观点，他们一般认为定性方法是自成一体的，可以完成从形成概念到检验假设的全部研究过程。在实际的应用研究中，定性方法和定量方法常常是交织在一起的，例如，克劳(Currall)等人在研究组织环境重要的群体过程时，通过内容分析把5年的参与观察资料量化，然后用统计分析来检验理论假定。格雷(Gray)和邓斯坦(Densten)在研究企业的控制能力时，利用潜变量模型把定性方法和定量方法有机结合在一起。雅各布斯(Jacobs)等人在研究比利时的家庭形态对配偶的家庭劳动分工影响时，首先用定量方法对纵向调查数据进行分析，从定量分析的结果中，又延伸出对核心概念的定性研究。这三个研究分别代表了定量和定性方法相互融合的三个方向：①克劳等人的研究代表着定性方法的实践者试图将定性数据尽可能量化的取向，近年来涌现出的处理调查数据中开放题器的编码问题的工具软件(如Words at，Smarttext等，注意：它们都是由著名的统计软件公司出品的处理定性资料的软件)，处理定性资料的传统内容分析软件(如Nvivo、MaxQDA、Kwalitan等)也开始提供将定性资料转换到常用统计软件的数据接口，这些工具上的革新将加快这种趋势的发展。②格雷和邓斯坦的工作代表了"方法论多元论"的取向，即在应用研究过程中，通过核心概念的测量模型，把定性研究和定量研究结合在一起。③雅各布斯等人的工作则代表了一部分定量研究者对过度形式化的定量方法的不满，并试图通过定性方法加以弥补。在定量研究领域中，对"模型设定"问题的关注，是定量方法重新试图返回定性研究这种取向的另外一种表现。

与社会调查技术和社会研究方法突飞猛进的现实相比，我国学术界在这些方面的论著的出版似乎显得有些迟缓。虽然已经翻译了美国的一小部分经典定量分析教材，如布莱洛克(Blalock)和巴比(Babie)的教材，也有自己编写的一些教材，如袁方等人的《社会研究原理和方法》、卢淑华的《社会统计学》等，此外，偏重软件操作的还有郭志刚的《社会统计分析方法——spss软件应用》、郭志刚的《logistic回归模型——方法与应用》、阮桂海的《spss for windows高级应用教程》等。在《社会学研究》等专业杂志上，也常常有一些定量分析的应用研究，可是专门的方法和应用模型研究却没有，也没有专门的方法研究期刊。仅就定量研究方法的介绍而言，也存在一些缺陷，主要表现在：

1. 原理和操作脱节。
2. 过分依赖某些商业软件,不全面。
3. 与中国的实证研究相脱节。
4. 不能反映当前方法研究的最新进展。

与定量研究方法相比,由于各种原因,定性研究方法的引进和介绍都比较少。在福特基金会资助的方法高级研讨班上,曾讨论过一些定性研究方法。在定性方法研究方面也有少数专著,如袁方和王汉生1997年出版的教程,陈向明2000年出版的专著。但总体说来,我们对定性研究方法还停留在初步介绍的阶段,主要的介绍也局限在定性研究的研究设计和资料收集的阶段上,对定性分析方法的介绍,则没有能够反映出当代定性方法的最新进展。特别是在定性分析工具(定性分析软件)的引进和研究上,基本上还是一个空白。虽然不乏一些出色的定性研究报告,但从方法研究上讲,我们才刚刚起步。当然,我们同时还应该注意到,在历史学领域,我国对定性资料的鉴别、考据和分析,积累了大量的经验和知识,这也应当是定性方法研究的知识来源之一,应努力发扬光大。

令人欣慰的是,社会研究方法的引进和出版方面相对滞后的状况终于有所改观。重庆大学出版社的编辑,以独到的学术眼光,逆当前出版界唯利是图的不良选题风气,投入了大量的人力、物力,组织出版"万卷方法"。自2004年至今,已引进社会科学研究方法方面的专著十余种,在我国社会科学界已经引起了一定的反响。然而,更为可贵的是,重庆大学出版社并未以已经取得的成绩而自满,而是再接再励,在原有"万卷方法"的基础上,进一步组织出版"万卷方法—社会科学研究方法经典译丛"。按我们的设想,"译丛"应该是一个开放的体系,旨在跟踪社会科学研究方法发展的前沿,引进和介绍这一方面的经典著作和最新成果。

"译丛"第一批有《抽样调查设计导论》《社会科学研究设计原理》《社会科学研究测量原理》《社会科学研究分析技术》《问卷设计手册》《回归分析法》《数据再分析法》《抽样调查设计导论》《社会网络分析法》《广义潜变量模型》《定性变量数据分析》和《复杂调查设计和分析方法》(书名也许有变化)等十余种,几乎囊括了研究设计、测量和分析方法的所有领域,涵盖从基础的回归分析到最前沿的潜变量分析和多水平模型等各种分析方法。无论是社会科学各专业的本科生、研究生,还是社会科学研究的学者都将从中有所收获。

"译丛"由中国社会科学院社会学所社会调查与方法研究室的多位研究人员担纲,主译者都是在社会研究方法各个领域中具有相当造诣的教师和研究人员。"译丛"的译者不仅仅把翻译看做是一个"翻译",而且也把它看做是一次再学习和再创新。

我们期待"译丛"的出版能对社会研究方法的研究、应用和教学有所推动。

沈崇麟　夏传玲

2010年12月于中国社科院社会学所社会调查与方法研究室

# 前 言

本书的主要目标是在尽可能广的范围上统一及扩展潜变量建模。本书所涵盖的模型包括多层次模型、纵贯性模型以及结构方程模型，另外还包括与它们相关的模型，如广义线性混合模型、随机系数模型、题器响应模型、因子模型、面板模型、重复测量模型、潜类模型及脆弱模型。

潜变量模型用于绝大多数经验研究学科当中，尽管常常被其他的名称所代替。本着本丛书的题目——“跨学科统计学丛书”，我们试图综合不同学科的视角，并在统计学、生物统计学、计量心理学以及计量经济学这几个学科的语言中进行转换（尽管我们并未声称掌握了所有这些学科的语言）。

我们认为模型的进展被这种“地方”行话所妨碍，并产生了隔阂。例如，计量经济学家和生物统计学家之间很少浏览对方的学术期刊。更加令人惊讶的还有学科内的部落化，这一点从《计量心理学》期刊上题器响应模型和因子模型之间缺乏交叉引用可以反映出来（尽管在同一个期刊上）。缺乏沟通的其中一个有害后果就是无法知晓对方学科所取得的有用进展，直到它们被“翻译”过来并发表在“适当的”文献上。例如，关于退出的模型（在社会科学中称为损耗）在过去的几十年中已经在计量经济学中取得了突出的进展，然而直到最近，它们才在统计学文献中被“发现”。

本书由两个部分所组成：方法论与应用研究。在第 1 章，我们讨论潜变量的概念、用途以及解释。在第 2 章，我们将不同学科所使用的针对不同响应类型的模型汇总在一起。在第 3 章对经典潜变量模型进行了回顾之后，我们在第 4 章将第 2 章所考察过的所有响应类型的模型进行统一及扩展。已有的以及最新的关于模型辨识、模型估计、潜变量预测以及模型诊断的方法将详尽地在第 5 章到第 8 章进行阐释。

在应用研究部分的第 9 章到第 14 章，我们使用第 1 部分所发展的方法来分析来自生物学、医学、心理学、教育学、社会学、政治学、经济学、市场营销以及其他领域的问题。所有的应用研究都使用真实数据，但出于教学原因，我们的分析通常是简化了的。在所有这些应用中，我们都使用了与安德鲁·皮科尔斯合作开发的 Stata 程序包 gllamm。

我们希望本书的两个部分中丰富的交叉索引可以让读者既容易在应用章节找到方法论的例子（忽略前半部分），也容易在方法论章节中找到应用研究的统计学背景（忽略后半部分）。

前面的 3 章多内容是作为对贯穿本书的建模视角的相对温和的介绍。后续的方法论章节具有更强的技术性，部分原因也在于本框架的通用性。然而，我们也同样考虑更简单的特殊例子，使得标记不那么复杂，想法更加明晰，结果更加符合直觉。对于潜变量模型的解释及应用更加感兴趣的读者可以忽略第 4 章到第 8 章的大多数内容，他们可关注于应用章节部分。

本书是过去4年中我们在合作开发“广义线性潜变量及混合模型(GLLAMM)”及其软件的一其中个成果。我们感谢我们的合作者安德鲁·皮科尔斯所做的贡献。安德斯感谢他的“老板”——皮埃·马格纳斯,索菲亚感谢她的前“老板”布莱恩·艾弗瑞特,是他们给予了支持并且接受本书的工作优先于其他项目。布莱恩·艾弗瑞特、莱昂纳多·格里利、卡拉·拉姆皮奇尼以及另外两个匿名评审员阅读了本书稿,并给予我们许多有用的建议。另外,我们还要感谢艾瑞特·艾特金、比尔·布朗尼、斯蒂芬·詹金斯、安德鲁·皮克尔斯、斯文·奥韦·萨缪尔森以及杰罗恩·沃芒所给予的建设性评论。

大卫·克雷顿、皮埃·卡拉格·安德森、安东尼·希思、安德鲁·皮克尔斯以及本特·特拉恩为本书的应用研究部分提供了数据。我们感谢帕特里克·赫加蒂、BUGS 项目组、穆森夫妇、应用计量经济学杂志以及皇家统计学会将数据发布到互联网上以供访问。在海琳·罗肖夫的帮助下,我们还使用了来自英国数据中心以及挪威社会科学数据服务中心的数据。感谢查普曼-霍尔出版社的杰斯敏·奈姆和克斯蒂·斯特劳德,是他们的频繁而温和的提示保证本书取得稳固的进展。我们还感谢 LATEX 的开发者提供了这个不可估价的免费工具帮助我们准备书稿。

在每一个章节中,我们都共同写作,为本书做出几乎相同的贡献。尽管本书的写作是一项非常辛苦的工作,但我们在此过程中也享受了很多乐趣!

gllamm 软件、文档等可从http://www.gllamm.org 下载。本书应用研究中的部分数据和脚本可从http://www.gllamm.org/books 下载。

安德斯·斯科隆多
索菲亚·拉贝-赫斯基思

# 目 录

# 第 1 部分　方法论

# 1 潜变量无所不在

## 1.1 介绍

既然本书的核心内容是潜变量模型,因此很自然地从讨论"潜变量"这一概念的含义入手。从当前的情况看,潜变量已经以不同的方式被界定,其中一些将会在本章进行简要描述,尽管我们通常发现这些定义过于狭隘(参见 Bollen,2002)。在本书里,我们将一个"潜变量"简单地定义为这样一种随机变量,即其实现过程对我们来说是隐蔽的那些变量。这与显变量(manifest variables)正好相反,显变量的实现是可以观测到的。

关于潜变量建模(latent variable modeling)的怀疑论与偏见在统计学家中并不罕见。潜变量建模经常被认为是一种不确定的应用,充满着无法证实的假设以及未经检验的因果关系推论。这种看法至少可以从以下 3 点进行反驳:第一,任何合理的统计方法都可能被未经检验的模型设定(model specifications)和过度热衷的解释而滥用。第二,忽视潜变量通常意味着更强烈的假定,而不是将它们包括在内。潜变量建模因此可能被认为是对于排斥潜变量的简单分析的灵敏度分析。第三,潜变量建模中的很多假设能够进行经验的评估,当然也有一些假设是毫无必要的,我们将会在随后的章节中看到这一点。

此外,潜变量建模通常被认为是统计学中一块相当模糊的领域,仅仅局限于计量心理学。然而,潜变量现已遍及现代主流的统计学,并且广泛地应用于不同的学科中,如医学、经济学、工程学、心理学、地理学、营销学与生物学。潜变量的这种"无所不在"一般并没有被认识到,或许可能是因为潜变量在不同的文章中被给予不同的名字,如随机效应(random effect)、公共因子(common factor)和潜类别(latent class)等。

在本章中,我们将要论证潜变量可以用来描述以下现象:

- 带有测量误差的"真实"变量。
- 假设构念(Hypothetical constructs)。
- 未观测到的异质性。
- 数据缺失。
- 反事实性(Counterfactuals)或"潜在结果(potential outcomes)"。
- 类别变量(categorical variables)背后的潜响应。

潜变量还应用于:

- 产生灵活性的多元分布。
- 统合来源于不同的个体单元信息。

在本章中,我们对这些潜变量的应用都进行简要的介绍,在本书的方法论和应用研究将会对涉及的相关部分进行详细的解释和举例。

## 1.2 带有测量误差的"真实"变量

潜变量能够表述一个连续变量误差测量的"真实得分(true scores)"。对于单元$j$的连续项或测量值$y_j$,根据传统检验理论(如 Lord and Novick,1968)的测量模型可写为

$$y_j = \eta_j + \epsilon_j$$

其中,$\eta_j$是真实得分,$\epsilon_j$是测量误差。这些测量误差的均值为零,彼此互不相关且与真实得分也不相关。真实得分被定义为测量变量或目标项的期望值:

$$\eta_j = E(y_j)$$

该期望值建立在想象的重复实验之上(Lord and Novick,1968)。这就是潜变量的**期望值定义**。

通常,对于每个单元都会运用到若干个测量值$y_{ij}, i=1,\cdots,n_j$,传统测量模型则变为(可参阅 3.3.1 节)

$$y_{ij} = \eta_j + \epsilon_{ij}$$

图 1.1 给出一个$n_j=5$的测量的路径图。根据路径图的传统,圆环部分代表潜变量$\eta_j$,各个小矩形代表观测到的测量值$y_{ij}$,各个箭头代表线性关系(在这里回归系数设定为 1)。"单元$j$"标记意味着方框内的所有变量在各个单元之间变动,并且下标$j$在以后的图表中将省略不标。

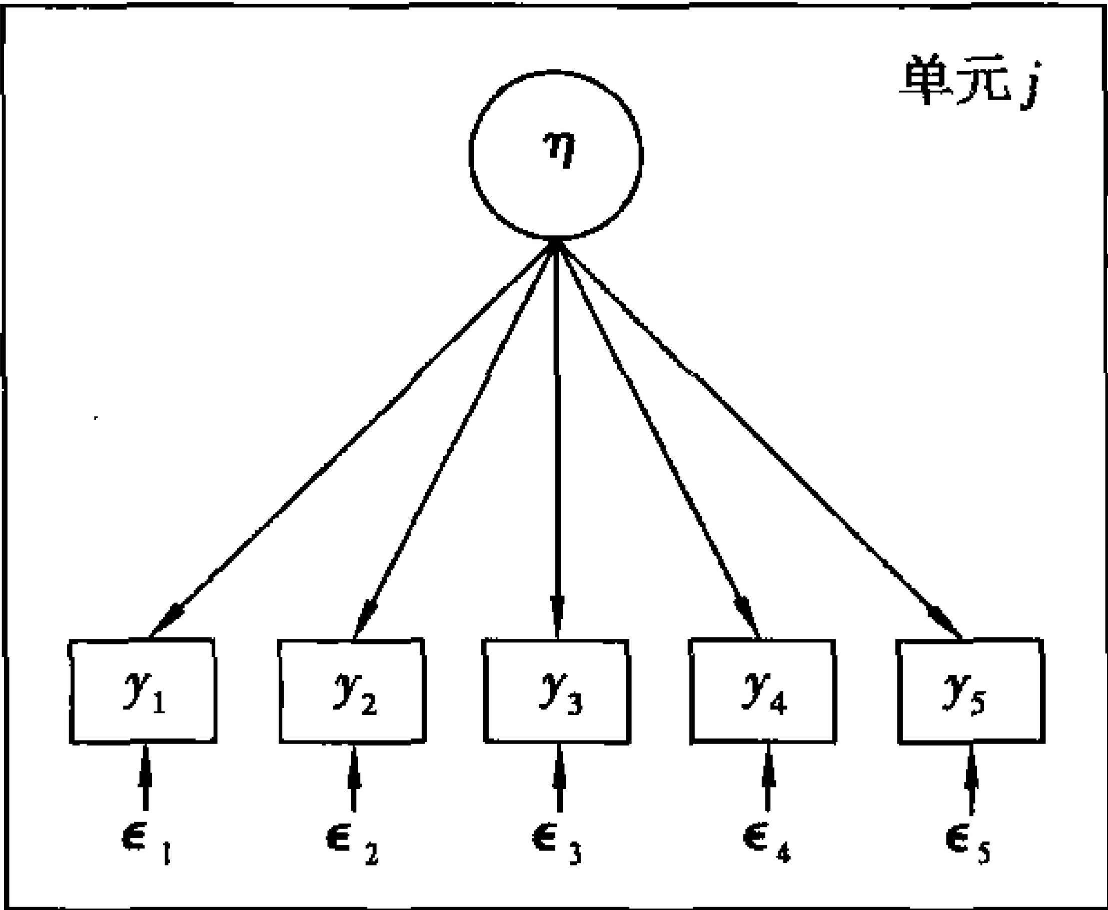



**图 1.1 传统测量模型的路径图**

测量模型通常用连续潜变量$\eta_j$来进行设定。当观测到的测量是连续变量时,这种模型被称为因子模型(参见 1.3 节式(1.1)),而在测量是类别变量时则称为题器响应模型(item response models)。在 3.3 节将对因子模型与题器响应模型进行更为详细的讨论。

有时,真实变量需要被解释为类别变量,一个典型的例子是医疗诊断(如有病或没病)。在这个案例中,测量通常也是分类的,与真实变量具有相同个数的类别。测量误差这时表现为错误分类(misclassification)。包含类别潜变量与测量变量(measured variables)的测量模型就是为人熟知的潜类模型(latent class models),在 3.4 节将进行论述。在 9.3 节将介绍其在一项心肌梗塞(心脏疾病)诊断中的应用。

测量模型的一个基本假设是(既针对连续变量也针对类别变量),测量是条件独立于给定潜变量的,也就是说,测量之间的相关性完全由于它们与潜变量的共同关联(common association)所造成的。这反映在图 1.1 的路径图中,即没有箭头直接将观测变量(observed variables)相连接。这种条件的或"局部的"独立性是潜变量的**局部独立定义**的

基础(如 Lord,1953;Lazarsfeld,1959)。

测量建模可以用来评估测量质量(measurement quality)。如果真实变量是连续的,测量质量通常根据个体测量的信度进行评估,参见3.3节。14.2节将介绍一项关于粗膳食纤维摄入研究的应用,其他应用实例在10.3节和10.4节都有介绍。如果真实变量是分类的,测量质量则一般根据分类错误率、灵敏度和特异性(specificity)表达出来。这在9.3节关于心肌梗塞诊断的应用中将进行探讨(同时可参见13.5节)。

当协变量的测量包含误差时(如 Carroll et al.,1995a;参见3.5节),测量模型还可以与回归模型相结合,以避免**回归稀释**(regression dilution)。在14.2节,我们讨论饮食中粗纤维摄入量对冠心病的效应模型,在这个例子中,粗纤维摄入的测量包含了误差,同时其中一个子群体可进行重复测量。有时,对于一个"校验样本(validation sample)"存在一个"黄金标准"测量,而对全部样本来说测量非常不准确。在校验样本中,给出一个与协变量缺失模型具有同样形式的模型,协变量的真实值因此是可以观测到的(尽管它是通过校验样本中的一个潜变量表现出来的),同时参见1.5节。我们将在14.3节讨论一个关于子宫颈癌病例对照研究的这种模型。

协变量测量误差模型经常需要相对严格的假设,诸如在给定真实值的条件下,条件独立于测量,非差异性测量误差(nondifferential measurement error)(在给定真实协变量的条件下,所测量的协变量条件独立于响应变量),正态分布的测量误差以及正态分布的真实协变量。这些假设大多能够被评估和或释放。例如,我们将要在14.3节中释放非差异性测量误差假设。在14.2节,当可进行重复测量时,我们运用非参数最大似然估计,从而无须设定真实协变量的分布。此外,忽略测量误差的"轻率的"分析,可能(但不一定!)比一个考虑到测量误差但是设定错误的模型产生更大的偏误。后者可以视为对前者的灵敏度分析。

## 1.3 假设构念

与含误差测量的真实变量——它们是假定为存在的(本体论的)——形成对比,假设构念拥有一个排他性的认识论地位(如 Messick,1981)。将假设构念看做是真实的,会因此引起一种具体化误差(reification error)。很多统计学家关于"潜变量"的怀疑论,也许可以归结于假设构念的形而上学性。另一方面,不依赖假设构念,沟通(communication)似乎是不可能的。如,一个"优秀统计学家"的概念是不真实的,但仍然是有用的,并且在统计学家中得到广泛的理解(尽管不容易界定)。

根据克朗巴哈(Cronbach,1971),一个构念是一个智慧的设计,依靠个人对事件的解析。这样,构念只不过是些概念。构念之间的关联提供关于观测关系的归纳性概要,作为详细描述理论规则之网络的一个基础(如 Cronbach and Meehl,1955)。南奈利和杜尔汉姆(Nunnally and Druham,1975:305)用以下方式对其进行了表达:

> ……科学家用来表示构念的这些词语,如,"焦虑"和"智力",在可观察的世界中并没有实际的对应物;它们仅仅是具有启发性的设计,用来揭示可观察的事物。

既然假设构念并不对应实际的现象,因而即使在原则上,它们也不能被直接地进行测量(如 Torgerson,1958;Goldberger,1972)。相反,这种构念是根据许多题器或间接"指标",被操作性地进行界定的,诸如一项智力测验中的回答。潜构念与观测指标之间的关联通常运用一个公共因子模型进行建模(Spearman,1904):

$$y_{ij} = \lambda_i \eta_j + \epsilon_{ij} \tag{1.1}$$

其中,$\eta_j$ 是潜变量或"公共因子",即表现为假设构念,$\lambda_i$ 是第 $i$ 个题器上的一个因子负荷,$\epsilon_{ij}$是一个独特因子(unique factor),表示题器 $i$ 的特殊性质和测量误差,可参见3.3.2节。这一因子模型可以用与传统测量模型同样的路径图(见图 1.1)进行描述,但从因子到指标的路径现在可以用因子负荷标示出来。

假设构念在心理学的研究中非常著名。事实上,可以公平地说,心理学以及类似学科中的大多数研究都涉及假设构念,诸如"自信""个性"和"生活满意度"(参见 10.4节)。社会学家经常涉及诸如"期望""异化"之类的构念,而政治科学家则对"政治功效"抱有相当的兴趣(参见 10.3 节)。在教育学中,研究者对诸如"算术能力"之类的构念感兴趣(参见 9.4 节)。必须指出的是,假设构念同时还应用于经济学等"硬"科学中,用以表示"固定收入"(如 Goldberger,1971)与"预期"(如 Griliches,1974)等概念。假设构念在医学中也是相当重要的,实例包括"抑郁症"(如 Dunn et al. ,1993)与"生活质量"(如 Fayers and Hand,2002)等。

到目前为止,我们已经讨论了连续的假设构念,它们似乎是最为普遍的情况。然而,有时考虑分类构念或类型也是正常的。社会学中一个明显的例子是"社会阶层"(如 Marx,1970)。在心理学中,"改变阶段(stages of change)"(无意图期、计划期、准备期、行动期、维持期和终止期)被认为是非常有用的,用于评估病人在他们向健康行为改变的"旅程"——如试图戒烟——中处于哪一阶段(如 Prochaska and DiClemente,1983)。在商业中,它是普遍的实践,用以将消费者分类进各种"市场区间",或者为目标营销,或者为服装产品。如,马吉德松和沃芒(Magidson and Vermunt,2002)运用潜类模型将银行消费者区分为四种类型:"寻求价值者""保守的储蓄者""主流者"与"投资者"。我们在 13.6节分析了一项关于咖啡生产者的市场细分的应用。在医学中,功能性综合征——诸如急性肠道综合征,它们是由一系列症状(其引发的原因是未知的)加以标定的,可以被视为分类型假设构念。这里,确定的症状具有相当高的同时出现的可能性这一事实,被用来作为一种标志,即它们可能是由同样的紊乱所造成的。

与在理论基础上界定假设构念相反,它们有时"起源"于一种探索性的分析,经典的实例是当潜变量被认为是连续时探索性因子分析的运用。分类潜变量的另一种类似情况是,运用探索性的潜类分析去推导出分类构念或类型。以这种方式发展理论的危险,已经由阿姆斯特朗(Armstrong,1967)进行了鲜明的论证。他在一个潜在因子(underlying factors)已被掌握的实例中运用探索性因子分析,这一潜在模型非常简单并且与数据完美拟合。尽管这个探索性因子分析"解释"了总体方差中的大部分,但它未能有效地获得这些已知的因子。在此,我们引用阿姆斯特朗所做的总结:

> 进行因子分析研究的成本在近些年已经得到实质性的下降。与早期相比,现在进行因子分析要比决定对什么进行因子分析容易得多。文献的大量增加能否将我们引领到更好的理论发展中,这还不太清楚。
>
> 因子分析可能提供了一种评价理论或提出理论修正的方法。然而,这要求将理论明确地指定为优先于数据的分析。否则,评价结论的标准是不充分的。如果大多数研究都是用来提出假设,而没有一个清晰的先验分析(priori analysis),我们的世界不久就会出现假设泛滥成灾的现象。

的确,细读当代的心理学杂志,就会清楚地发现他的预言已经得到应验;这部分世界已经充斥了各种假设。

作为一个具体的例子,阿姆斯特朗提到卡特尔(Cattell,1949)进行的一项试图揭示文化主要维度的研究。卡特尔获得的 12 项基本因子都是相当难以理解的,包括诸如"文明

丰富”“富于创见的勤奋”“资产阶级庸俗主义”之类抽象的东西。然而,类似这种有问题的应用仍然大量存在于心理学当中。新近的一个突出范例是人格心理学中流行的“大五人格理论”(big-five theory)(如 Costa and McRae,1985),强烈地提倡人格是由以下5个维度来标定的,即“外向性”“随和性”“尽责性”“神经质(neurotisism)”与“开放性”。这一“理论”在很大程度上是来源于探索性因子分析,但这些因子仍然被给予了本体论的地位(被认为是真实存在的)。瓦斯恩德和斯科隆多(Vassend and Skrondal,1995,1997,2004)对大五人格理论持批评态度,认为人格的传统分析工具充满了概念上和统计上的问题。

尽管连续的假设构念通常通过公共因子进行建模,但情况并非总是如此。几种其他的多元统计方法也已经应用于探索潜在数据(underlying data)的“维度”。实例包括主成分分析(如 Joliffe,1986)、偏最小二乘法(partial least squares,PLS)(如 Lohmöller,1989)与多维尺度分析(multidimensional scaling)(如 Kruskal and Wish,1978)。分类“构念”可以通过聚类分析、有限混合模型构建或多维尺度分析(如 Shepard,1974)推导出来。然而,我们并不将这些方法产生的“维度”或“组别”视为潜变量,因为它们仅仅代表了数据资料的转化形式或几何学特征,而不是一个统计模型中的要素。事实上,本特勒(Bentler,1982)将潜变量定义为:无法仅仅通过显变量的函数进行表达的变量。另一局限性是,这些方法通常是严格探索性的,不容许强加任何基于研究设计、前期研究或实体理论(substantive theory)的假设结构(hypothesized structure)以及对其进行检验。这同时也是探索性因子分析和潜类分析所面临的问题。不同类型的统计模型在8.2.1节进行了对比,不同的建模策略在8.2.2节进行讨论。

我们必须承认假设构念在很多学科中是非常有用的,研究**构念效度**(construct validity)的方法、测量方法以及构念之间的相互关系,都是十分重要的。本特勒(Bentler,1978)和巴格兹(Bagozzi,1980,1981)以及其他人一道,认为包含潜变量的结构方程模型是分析构念效度的最佳方法(参见3.5节)。这种明确地指定了假设结构的验证性方法(confirmatory approach),可以被视为传统构念效度概念的一种形式化(formalization)(如 Cronbach and Meehl,1955)。克朗巴哈和米赫尔将构念理解为卷入一种既是理论的关联又是经验的关联的网络中;克朗巴哈(Cronbach,1971:476)的表述如下:

> 将构念彼此联结的陈述,以及含构念的可观测指标,组成了一种法则性的网络(nomological network)(法则性的意思是像法律的、定律性的)。

几种类型的效度已经在心理学中进行了概括,包括“内容效度”“收敛效度”“同期效度”“判别效度”“预测效度”与“法则效度(nomological validity)”(American Psychological Association et al,1974)。构念效度已经成为效度的一种层级视角和统一视角的核心,能将这些类型整合起来(如 Silva,1993)。

潜变量建模的一个优点是,我们可以通过模型拟合(参见8.5节),或者以不同方式建立模型来检验假设结构的可成立性(tenability)。例如,**收敛效度**可以通过设定模型进行评估,在这一模型中,指标的设计是用来反映某一给定的构念,并仅仅反映了那一个构念而非其他。我们对图1.2中的上图进行一下描述,第一个因子通过1~3的题器进行测量,而第二个因子由4~6的题器进行测量。如果拒绝这个模型而支持下图中的模型,即其中题器5在两个因子上都有负荷,那么收敛效度将不再有效。在这个实例中,可以采取两种替代性的行动路线,或者调节下图中的题器,或者抛弃这个模型。后一种方法在题器响应模型中用得很普遍,可能会因为“自适应(self-fitting)”(如 Goldstein,1994)而受到批评,但如果理论建立得很好,则可能被证明是正确的。

**判别效度**可以通过检查构念的独特性(uniqueness)进行检验,在这个意义上,估计的

构念之间的相关性不应该太大。在图 1.2 的第二个的方框图中，估计协方差 $\hat{\psi}_{12}$ 较大，其判别效度是有疑问的。

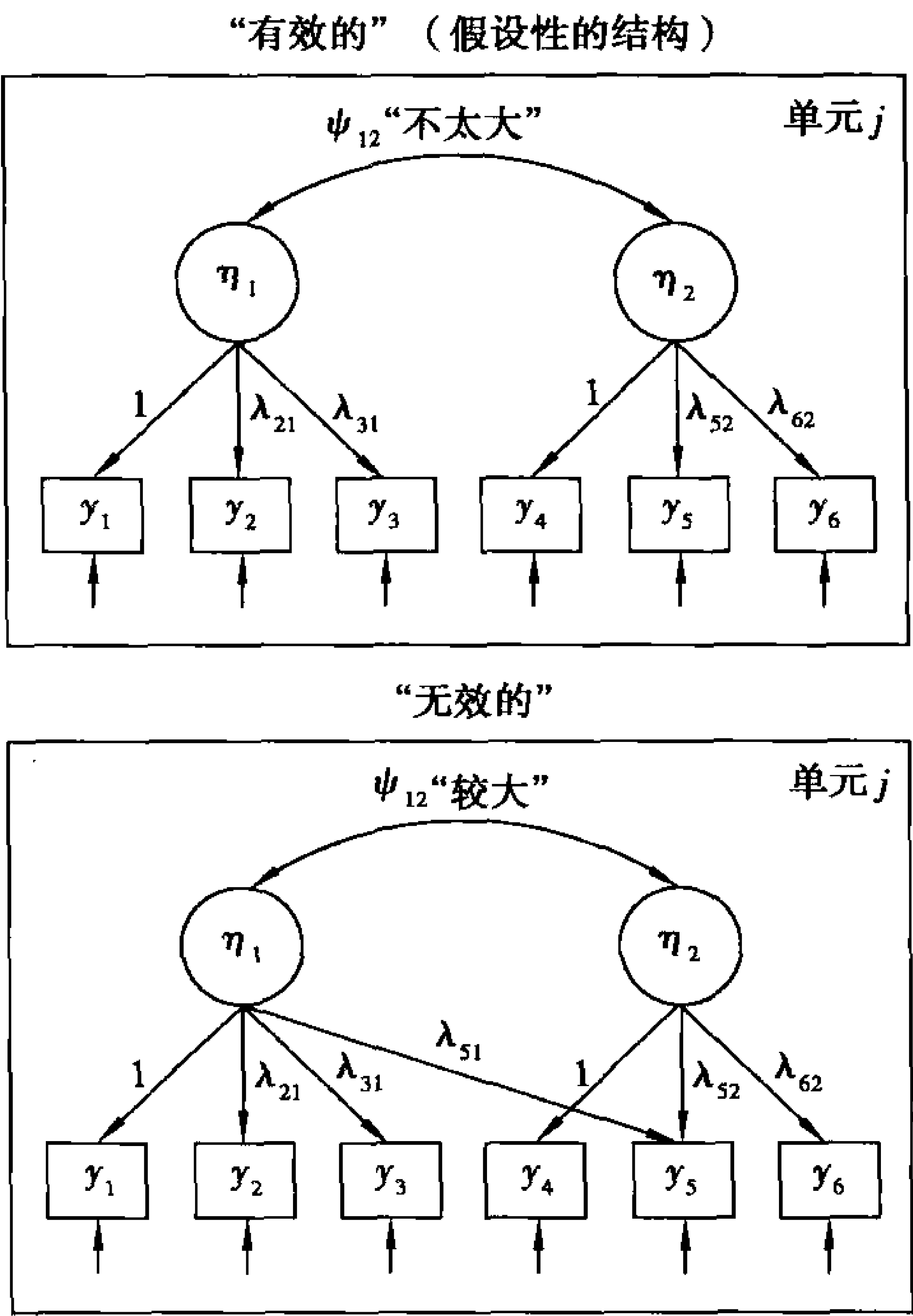



**图 1.2　收敛效度(convergent validity)与判别效度(discriminant validity)**

**法则效度**可以通过检验结构模型——该模型结合了归纳构念之间关系的理论——的可成立性来进行有效的评估。例如，考虑图 1.3 中左边方框的"完全中介性(complete mediation)"的模型(如 Baron and Kenny,1986)，假定 $\eta_1$(如"反常的双亲关系")对于 $\eta_3$(如"抑郁")没有直接的效应，而仅仅通过中介 $\eta_2$(如"个性")产生间接效应。如果拒绝这个模型，而支持一个 $\eta_1$ 对 $\eta_3$ 既有直接效应又有间接效应的模型——如图 1.3 所示的右边方框，那么这个完全中介性模型就不具有"法则"效度。

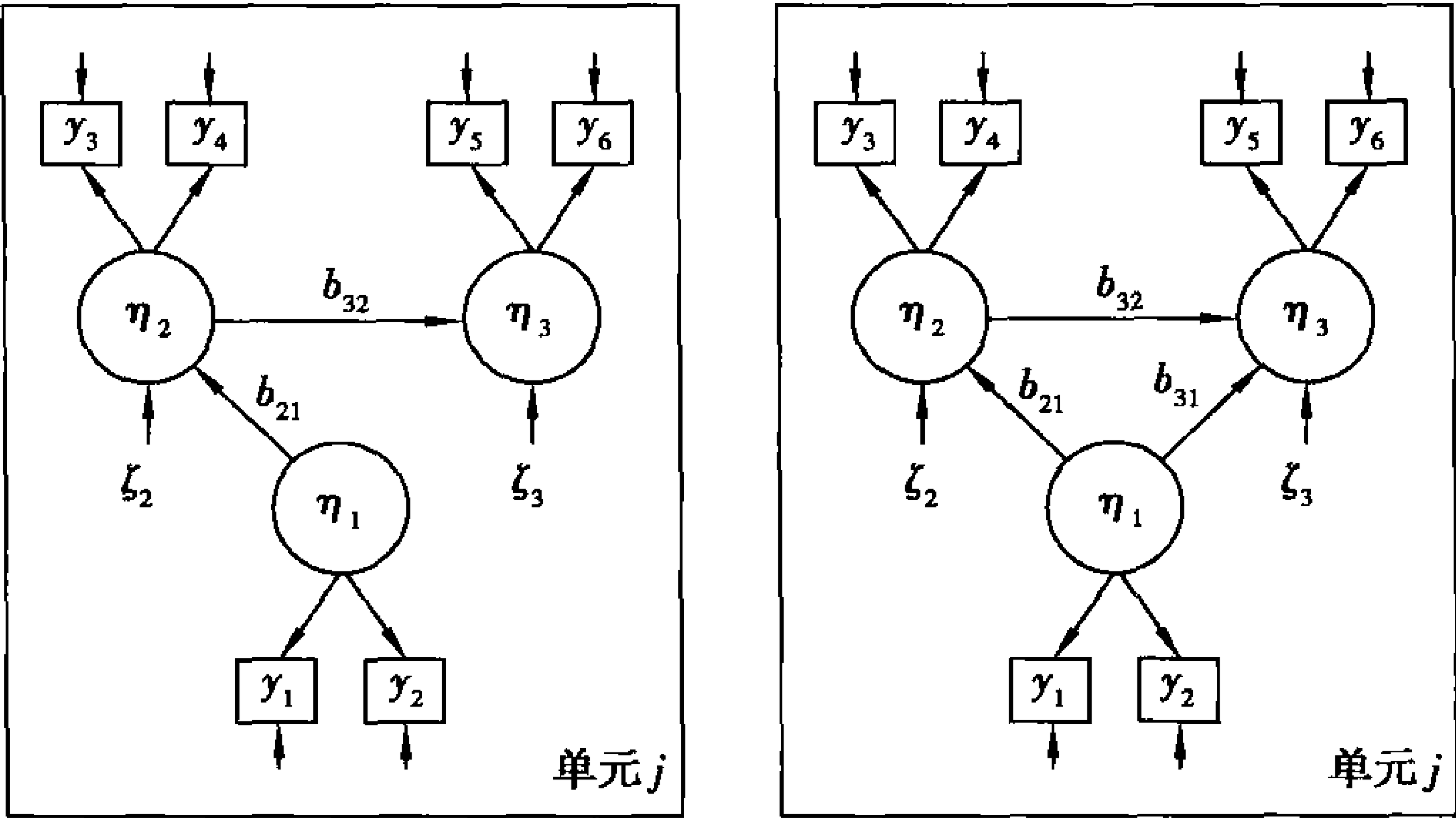



**图 1.3　法则效度**

我们可以参考巴格兹(Bagozzi,1981)的一个应用——检查态度的期望值和语义差异模型的构念效度。

## 1.4 未观测到的异质性(Unobserved heterogeneity)

统计模型的一个主要目的是根据可观测的协变量的变异性来“解释”响应变量的变异性,有时称为“观测到的异质性(observed heterogeneity)”。然而,在实践中,并非所有相关的协变量都是可以观测到的,这就会导致**未观测到的异质性**。纳入潜变量(在本文中通常指的是**随机效应**),在统计模型中是一种考虑未观测到的异质性的一般方式。

随机效应模型广泛地应用于多种问题。实例包括纵贯性分析(longitudinal analysis)(如 Laird and Ware,1982)、元分析(meta-analysis)(如 DerSimonian and Laird,1986)、捕获再捕获研究(capture-recapture studies)(如 Coull and Agresti,1999)、结合分析(conjoint analysis)(如 Creen and Srinivasan,1990)、生统遗传学(biometrical genetics)(如 Neale and Cardon,1992)及疾病制图(disease mapping)(如 Clayton and Kaldor,1987)。未观测到的异质性模型的应用将在后续章节讨论,其中,9.2 节和 11.3 节是纵贯性研究,9.5 节是元分析,9.7 节是捕获再捕获研究,11.4 节是小域估计(small area estimation)和疾病制图,13.6 节是营销中的结合分析。

需要注意的是,未观测到的异质性并非一种假设构念,因为它仅仅是指所有未观测到的协变量的混合效应,并且没有被赋予超出这点的任何意义。遗传学研究中的随机效应也许处于一种中间位置,因为它们被解释为共享的和非共享的基因效应和环境效应。

当这些单元聚集在一起时,共享的未观测到的异质性可能导致这些响应中的“组内”依赖性(“intra-cluster”dependence),甚至在对观测的协变量加进约束条件后。这在关于 10 个群组的图 1.4 中进行了阐述,每个群组含两个单元并且没有协变量。这里,异质性反映在这些群组均值(水平的短线所示)围绕总均值(横线所示)的分布上,从而导致组内相关性,因为一个群组的两个响应往往趋向于处在总体均值的同一方向。这一现象对于纵向数据或面板数据来说是普遍的,对同一单元的观测资料受到同一(共享的)特定单元(unit-specific)未观测到的异质性的效应。9.2 节介绍了一个涉及印度尼西亚儿童呼吸道感染的重复测量的例子。其他关于群组数据的实例包括家庭中的个体成员,或者学校中的儿童。

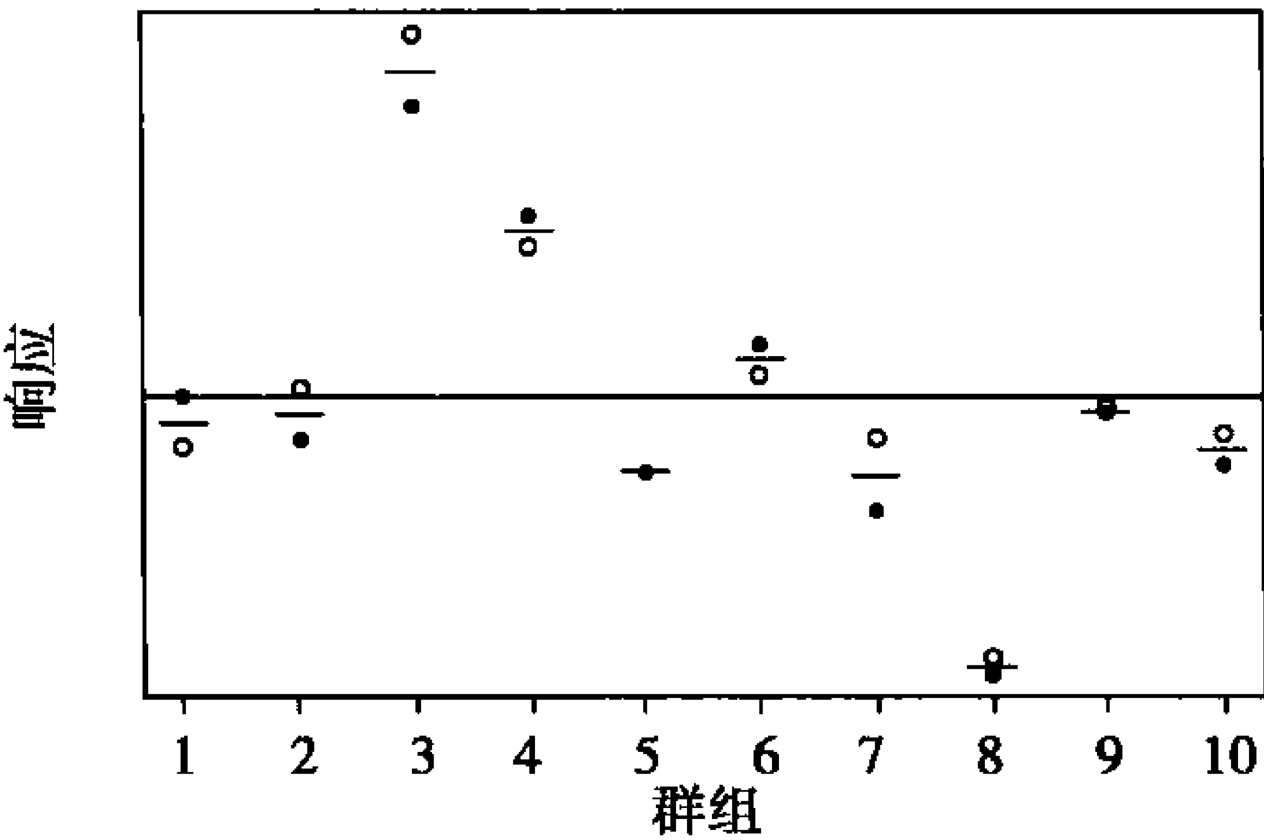



**图 1.4 组间异质性与组内相关性**

另外一个不同类型的群组数据是事件数数据,如一个癫痫患者一周发作的次数。尽管每一个个体只有一个数字,但每一个数字可能包含其发生很可能受到共享的特定单元协变量的效应的几个事件。在凭经验发现事件的“倾向”中,作为结果的变异性可能导致“过度离散(overdispersion)”,在泊松模型(Poisson model)中意味着方差要大于均值(参见 2.3.1 节与 11.2 节)。未观测到的变异性的这两种结果,即组内相关性与过度离散,如果不能做出正确的说明就会导致错误的推论。对未观测到的变异性进行说明的一种方式

是，在一个回归模型中纳入一个随机截距(random intercept)。就群组数据而言，同一群组中的单元必须共享相同的值或者随机效应的实现结果(realization)。在图 1.4 中，随机截距指的是群组均值(水平短条)与总均值的差异。

在多层次数据或层级数据中，通常存在几种群组层次，家庭中个体成员的面板数据就是一个例子。我们可以据此在每一个更高的层次上运用潜变量，以便在这一层次上描述未观测到的异质性。一个关于面板波动项(panel wave)$i$、个体$j$与家庭$k$的简单三层次随机截距回归模型可写为：

$$y_{ijk} = \eta_{0jk} + \beta_1 x_{ijk} + \epsilon_{ijk}$$
$$\eta_{0jk} = \gamma_{00} + \zeta_{jk}^{(2)} + \zeta_k^{(3)}$$

在$y_{ijk}$的第一层次模型中，$x_{ijk}$是"固定"回归系数为$\beta_1$的一项协变量，$\eta_{0jk}$是均值为$\gamma_{00}$、在第二层次和第三层次上残差分别为$\zeta_{jk}^{(2)}$和$\zeta_k^{(3)}$的一项随机截距。这一模型的随机部分(并非指$\beta_1 x_{ijk}$和$\gamma_{00}$)在图 1.5 中用路径图的形式进行表示，图 1.5 涉及拥有 3 名个体成员的家庭分别在 3 个事件、1 个事件和 2 个事件中的参与。多层次模型将在 3.2 节进行详细的讨论。

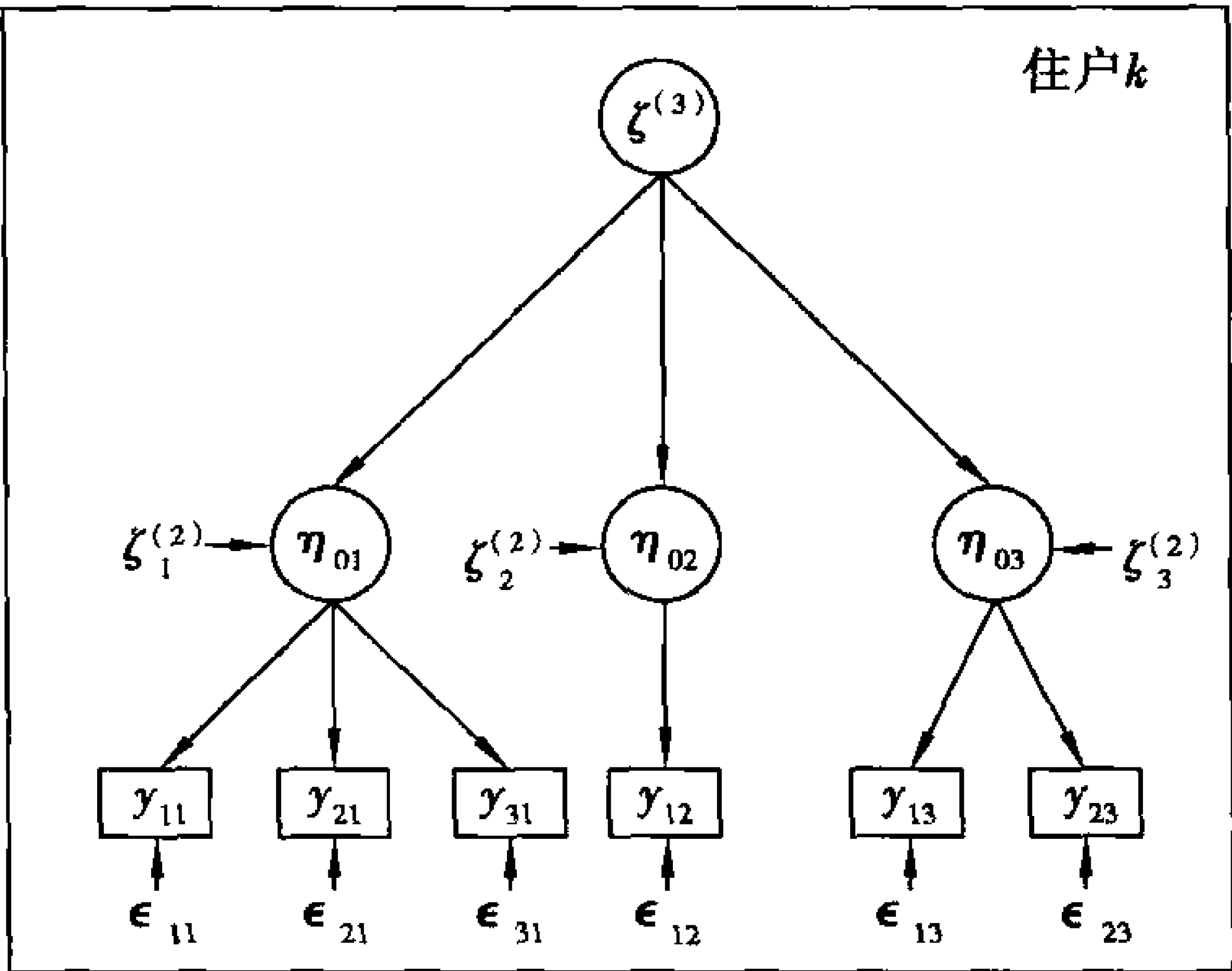



**图 1.5 三层次随机截距模型的路径图**

一个协变量对于响应的效应也可能在群组之间是不同的，这些群组可以通过包含特定群组的随机系数建立模型。如，癫痫病发作率随着时间的变动情况，可能会如 11.3 节所研究的那样，在各个治疗者之间是不同的。类似地，政治距离(党派与投票者之间)对于政党选择次序的效应在各个选区可能是多样化的，这是 13.4 节所讨论的内容。有时，未观测到的异质性是离散的，各个单元落在截然不同的群组中(参见 12.4 节的一项实际应用)。

未观测到的异质性有一个重要推论，即响应与未观测到的协变量之间的关系在单元(或群组)和总体层次上是不同的。一个突出的例子是生存分析(survival analysis)(如 Aalen，1988)中的弱点，其中，由于随后可能出现的不清楚的变异性，使总体层次上的风险(hazards)可能与单元层次上的风险完全不同。导致这种情况的原因是，某些个体比其他人更加"脆弱"，相对于用观测到的协变量进行解释，他们更容易受到事件的效应。这些个体可能倾向于在早期经历事件后，因为更少的脆弱而离开。因此，即使个体的风险是稳定的，总体平均的风险也会随着时间推移而降低。我们将在 12.4 节讨论心绞痛治疗的多元脆弱模型。

单元(或群组)和总体层次的效应之间的区别，对于二项响应来说也是十分重要的。如果多个特定单元模型都是这样的正态概率(probit)回归——具有不同截距(由于未观测到的异质性所造成的)但同时共享一个针对单一协变量的公共系数，那么总体平均模型就是一个具有衰减系数(attenuated coefficient)的正态概率回归。这在图 1.6 中进行了

阐述,其中粗体显示的总体平均曲线的倾斜度要小于特定单元曲线(参见 4.8.1 节的一个公式推导)。究竟是特定单元还是总体平均的效应具有重要性,将依赖于特定的情境。如,总体平均效应在公共卫生领域通常受到关注,因为公共卫生的焦点在于总体水平。另一方面,在一个临床环境里,对于个体患者及其医师来说,特定患者效应显然更为重要。更重要的,由于因果过程必须在单元层次而不是总体层次上得以实现,因此,关于因果关系的研究就需要特定单元的效应。如果存在对于单元的重复观测资料,特定单元效应就能够通过在模型中纳入随机效应而进行估计。

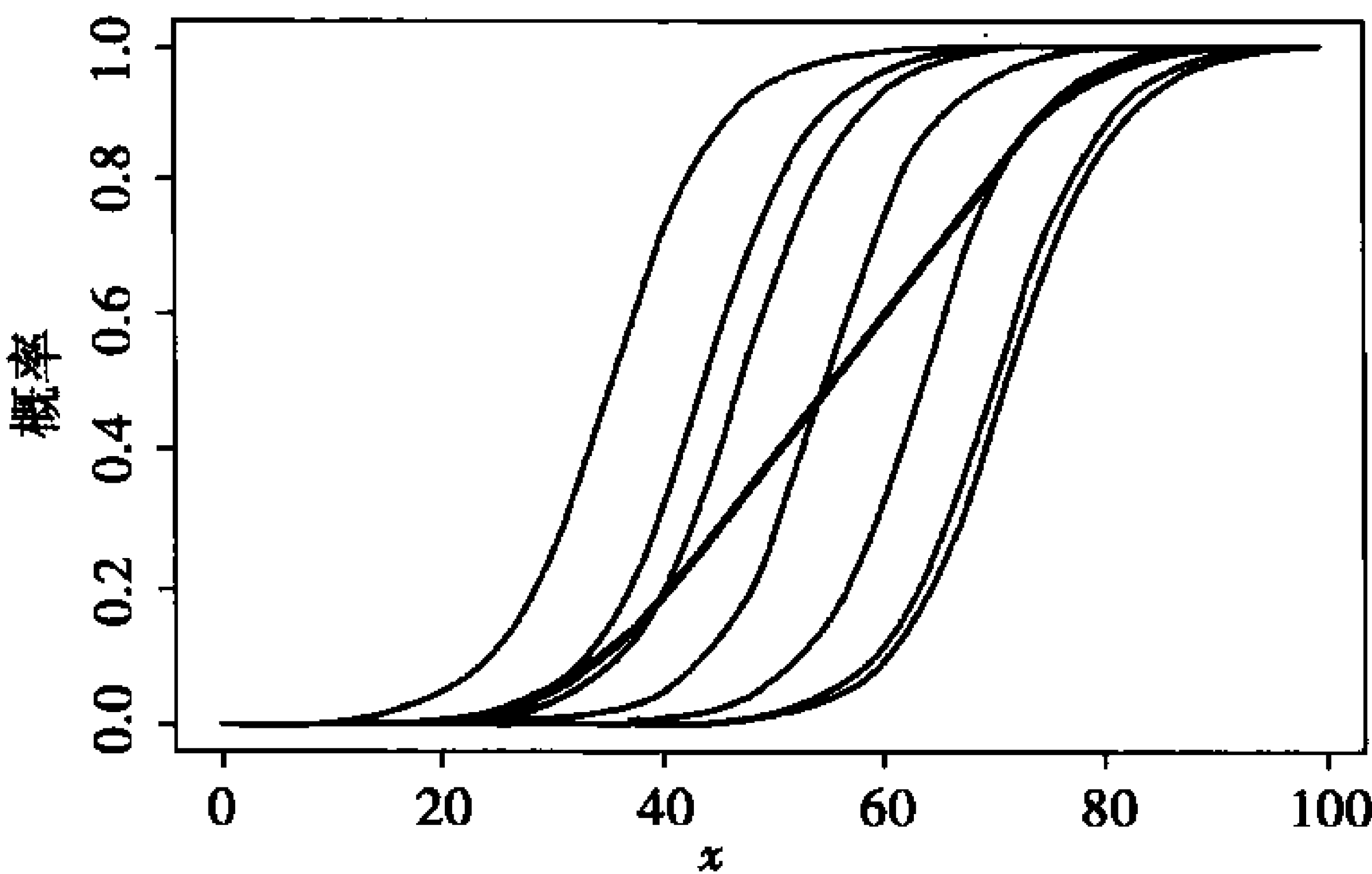



**图 1.6 特定单元正态概率回归与总体平均正态概率回归**

在前面的例子中,我们已经考虑了在多种事件中观测的单元,而现在我们将回到更为一般的情况,即单元(或事件)嵌套于群组(或者单元)当中的情况。注意这点是十分重要的:协变量与响应变量之间的关系在单元层次和群组层次上可能是迥然不同的。以群组层次上的聚集数据(aggregated data)为基础,对单元层次上的效应所作的推论,可能会由此导致所谓的"生态学谬误(ecological fallacy)"(如 Robinson,1950)。罗宾森的经典实例关注的是黑人的百分比与地区层次上的文盲之间的相关性,估计的结果是 0.95,这与黑人和个体文盲之间的个体层次的相关性估计有着明显不同,估计值仅仅是 0.20。

图 1.7 阐述了群组内效应(within-cluster effects)可能会与群组间效应非常不同,甚至是完全相反的方向。图中的群组可以代表不同国家,响应 $y$ 可以作为健康成果,比如寿命;而解释变量 $x$ 可以作为一种不健康享受的显露(exposure),比如红肉。在一个国家内,逐渐增加的不健康享受的显露是与逐渐恶化的健康(如图中带圆点的下斜线所反映的)相联系的。另一方面,在国家之间,渐增的平均显露是与逐渐增加的平均健康水平(如图中上斜的虚线所示)相联系的,因为它也是与不断提高的平均生活标准相联系。

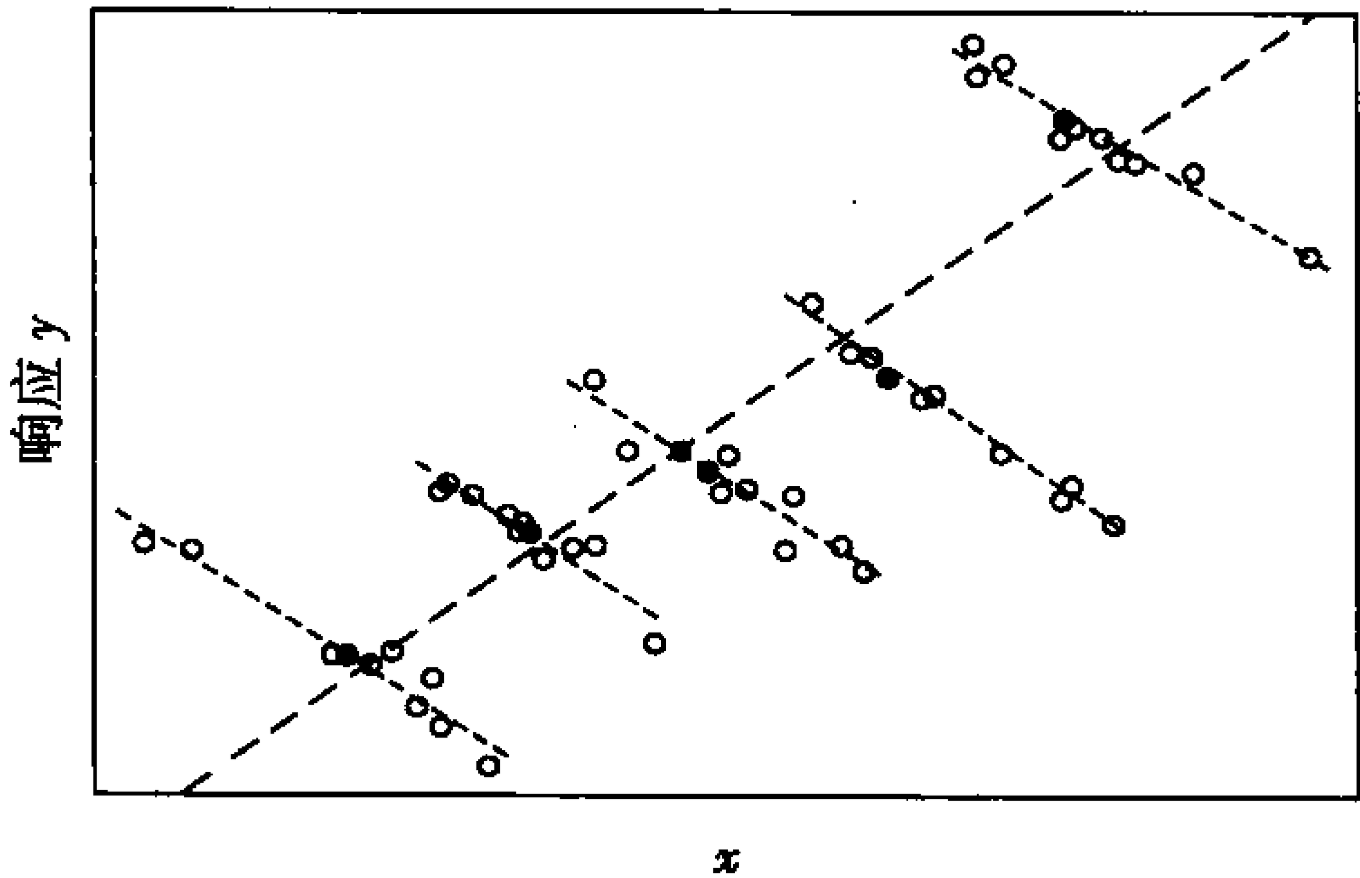



**图 1.7 群组内与群组间效应**

另一个重要范例是纵贯性数据,其中,单元内的下降可以代表一种年龄效应,但是单元间的提高可能是一种同期群效应(cohort effect)。如果仅仅能够获得截面数据(cross-sectional data),我们就无法区别年龄效应与同期群效应。例如,年长人群相比年轻人群

所表现出来的更强的保守性，可能是由于他们已经到了生命的一个更晚的阶段，或者是由于出生在一个不同的时代。我们进一步来看这个纵贯性模型

$$y_{ij} = \beta_0 + \beta_C x_{1j} + \beta_L(x_{ij} - x_{1j}) + \zeta_j + \epsilon_{ij}$$

其中，$y_{ij}$是单元$j$在事件$i$上的响应，$x_{ij}$可以是年龄，$\zeta_j$是一个特定对象的随机截距。这一纵向设计允许对横截面（或同期群）效应$\beta_C$和纵向效应$\beta_L$进行单独估计。这一特性是与3.2.2节涉及的随机效应与协变量之间的相关性问题密切相关的。

在随机效应模型中，最典型的是假设随机效应属于多元正态分布。模型诊断（参见8.6节）可以对这种假定进行评估，尽管它们可能并不十分灵敏。幸运的是，这些诊断推理在很多情况下对于随机效应分布的错误设定（misspecifications）具有相当稳健的作用（如Bartholomew，1988，1994）。此外，我们可以通过运用非参数最大似然估计，来释放这一分布假设（参见6.5节）。11.2节关于儿童患龋齿的模型、11.3节癫痫症发作率的模型以及14.2节食物与心脏病的模型都运用到这一方法。

## 1.5 缺失值与反事实

潜变量可以代表部分观测变量（即那些在单元子集上进行观测的变量）的缺失值。通常情况下，缺失值是被假定为已经“了解”但因为某些原因而没有记录。然而，缺失值有时只有在“反事实”推理的情况下才可能得到理解，如当一个协变量具有一个不同取值的时候。

如果一个**协变量**对于某些单元来说是缺失的，这些单元明显地不会有助于参数估计。这些单元的错失，导致了削减效率（reduced efficiency），这可以通过运用（广泛的）插值技术（imputation）填补缺失的协变量值而得到克服（如Rubin，1987；Schafer，1997）。然后这些单元就能为响应与其他协变量之间的关系贡献信息。作为对协变量值进行插值的替代方法，我们可以对插值模型与所研究的模型进行联合估计，从而整合针对缺失值的“插值分布”似然（imputation distribution likelihood）。在这里，缺失值可以通过一个潜变量进行描述，假定它们在所研究的模型中与观测值具有相同的分布并且与响应具有相同的关系（如回归参数）。一个协变量缺失问题的例子是，当存在一个校验样本（其中，真实协变量是观测到的）时的协变量测量误差（参见14.3节）。另一个例子是对不服从的随机干预中“服从者平均因果效应（complier average causal effects）”（Imbens and Rubin，1997b）的估计，其中，服从在控制组中没有进行观测（参见14.4节）。在这里，控制组中的服从情况也可以被视为“反事实”。

如果某些单元的**响应**是缺失的，只要至少存在一个观测响应，这些单元就可以有助于参数估计，如果数据是随机性缺失（MAR）（如Rubin，1976；Little and Rubin，2002），就可以产生一致性的参数估计（parameter estimates）。然而，如果响应并非随机性缺失（NMAR），忽略数据缺失的机制可能导致有偏的参数估计。这可以通过对真实过程与缺失过程的联合建模而得到处理。在“选择模型（selection model）”方法中（如Little，1995），关于缺失过程对未观测到响应的依赖性，有着明确的模型分析。

例如，在一项由于学生退学或生员减少而导致响应出现缺失值的纵贯性研究中，豪斯曼和怀斯（Hausman and Wise，1979）曾介绍了一个计量经济学中的模型，该模型后来由迪歌和肯沃德（Diggle and Kenward，1994）在统计文献中重新发现。在这里，每一时间点（假定它还没有发生）的退学学生运用一个使用先前响应和同期响应作为协变量的对数概率（或正态概率）回归模型进行建模。如果退学学生出现，同期响应没有被观测到，则

可以通过一个潜变量进行表示。于是这里的似然就是响应与学生退学情况(整合了这一潜变量之后的退学情况)的联合似然(joint likeihood)。在时间 $i$ 上,关于单元 $j$ 的所研究响应的模型 $y_{ij}$,其简单形式可写为

$$y_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \eta_j + \epsilon_{ij}$$

其中,$\eta_j$ 是一个特定单元的随机截距。退学学生变量可以构建如下模型,如果单元 $j$ 在时间 $i$ 上退学,则 $d_{ij}=1$,否则为 0:

$$\text{logit}[\Pr(d_{ij} = 1 \mid \mathbf{y}_j)] = \alpha_1 y_{ij} + \alpha_2 y_{i-1,j}$$

其中,当 $y_{ij}$是未观测到的时候,它由一个潜变量 $y^*_{ij}$ 代替。图 1.8 显示了一个具有跨越 3 个时间点的完全数据的单元的路径图,以及一个在时间点 2 上退学的单元的路径图。在后者中,时间点 2 上的响应由一个潜变量进行描述。应该注意的是,尽管它们并非线性的,但我们是用箭头来表示对数概率回归。这一分析可以直接被视为对 NMAR 缺失数据的纠正。然而,因为 $\alpha_1$ 的估计严重依赖于对 $y_{ij}$分布的假定,故它是可以用来代替将结果解释为一个 MAR 假定的灵敏度分析。

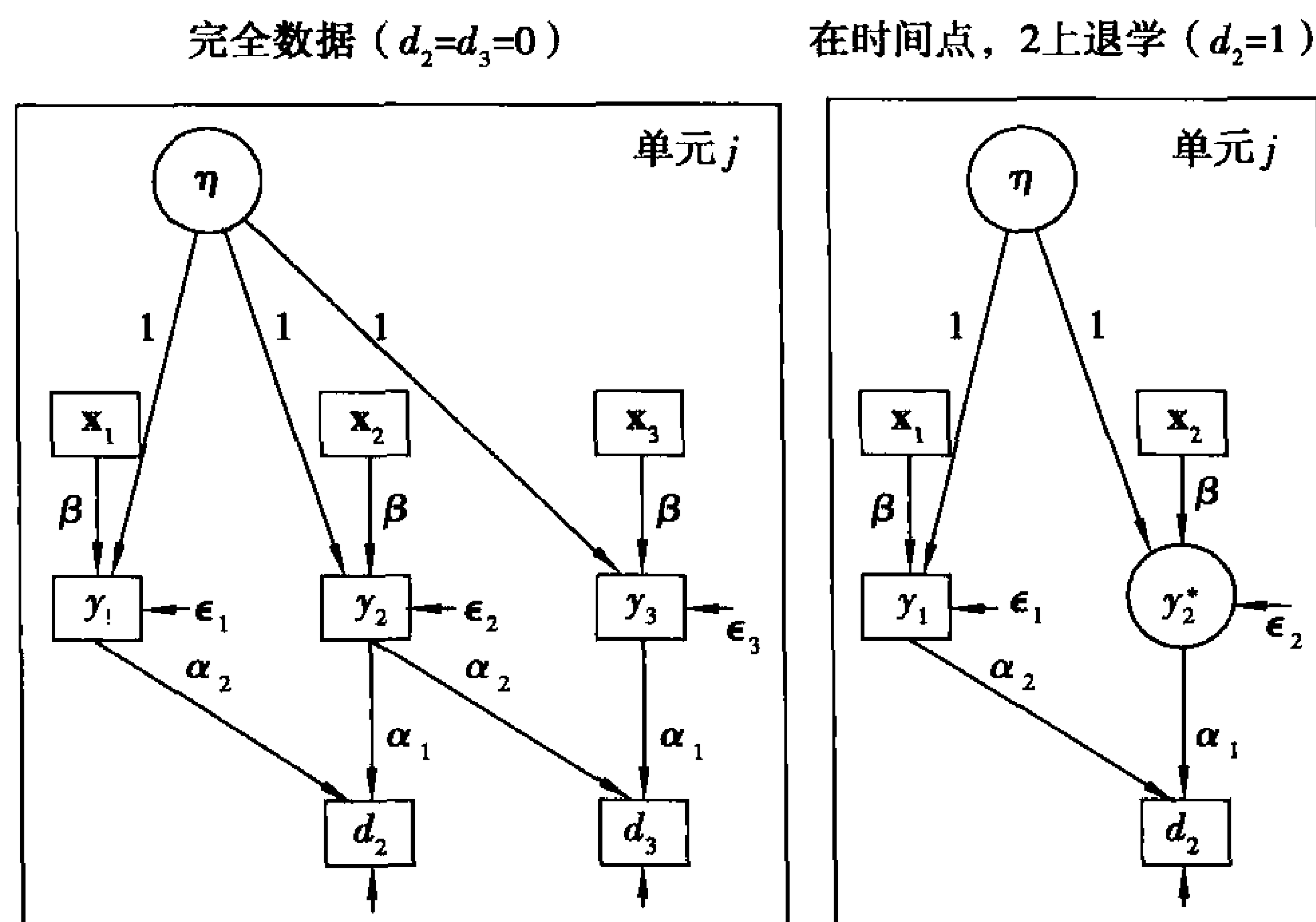



图 1.8 豪斯曼—怀斯—迪歌—肯沃德(Hausman-Wise-Diggle-Kenward)退学模型路径图

## 1.6 潜响应

潜变量可以用来表示这样一种连续变量,即可以观测到的不明显的、“粗化的(coarsened)”响应,如二项响应或定序响应(参见 2.4 节)。二项响应的潜响应解释由皮尔森(Pearson,1901)在正态分布的潜响应中进行了介绍。尽管潜响应方程最为普遍地应用于这样的正态概率模型,但它也适用于对数概率模型和余重对数模型(complementary log-log models)。在二项变量的情况下,观测响应 $y_i$ 可以通过结合关于不明显的/潜在的连续响应 $y^*_i$ 的回归模型,来建立模型:

$$y^*_i = \mathbf{x}'_i\boldsymbol{\beta} + \epsilon_i$$

以及一个阈值模型(threshold model):

$$y_i = \begin{cases} 1 & \text{当} \quad y^*_i > 0 \\ 0 & \text{其他} \end{cases}$$

如果 $\epsilon_i$ 是标准正态的,该模型就与正态概率模型相一致;如果 $\epsilon_i$ 服从对数分布,则相当于一个对数概率模型;而如果 $\epsilon_i$ 服从冈贝尔分布,则相当于一个余重对数模型。对于某一潜响应的一般性解释,是这种“倾向性”——即倾向于拥有一个正的观测响应 $y_i=1$。

在遗传学中,潜响应被理解为这样一种"倾向"——去发展一种特性或表现型,比如 I 型糖尿病(如 Falconer,1981)。赫克曼(Heckman,1978)对美国政府是否采用公正就业立法进行了模型分析,并且将相应的潜响应描述为"情感"偏好的公平就业立法。在毒理学中,一个单元对一种药品的"耐受性"是这个单元所能忍受的最大剂量,因此,超出这一剂量就会导致死亡(如 Finney,1971)。

在决策的情境中,个体会选择最优先的替代性方法,或者会依据偏好而形成一个替代性方法的排列,潜响应可以被认为是效用差异。如,考虑这一情景,即往返者必须在电车和公共汽车之间进行选择。电车与公共汽车的效用 $u_i^{car}$ 与 $u_i^{bus}$ 可能依赖于它们分别占用的往返者的交通时间以及其他协变量。如果 $u_i^{car} > u_i^{bus}$,或者是 $y_i^* = u_i^{car} - u_i^{bus} > 0$,往返者就会决定乘坐电车。这种相比较的响应与几种应用的模型都在第 13 章进行介绍。

潜响应的引入引起了很多批评,这并不奇怪。如,尤勒(Yule,1912:611-612)如此评论道:

> ……所有那些因天花而死亡的人都是同等地死亡:没有人比其他人多死一点或者少死一点,并且亡者与生存者完全不同。

皮尔森和海伦(Pearson and Heron,1913:159)对此回应道:

> ……如果尤勒先生的观点是可以接受的话,就会对现代统计理论的发展产生无法弥补的破坏。

暂且将这种哲学的(抑或个人的?)争论搁置一边,正态概率模型的潜响应方程毫无疑问是非常有用的,无论这一潜响应是否能被赋予一个真正含义。例如,在潜响应方程中,我们可以通过简单地允许这些潜响应是相互关联的四项相关(tetrachoric correlation),来为二项变量之间的相关性设定模型。类似地,我们可以通过允许潜响应与观测的连续响应相关("双序列相关(biserial correlations)"),来对二项变量与连续变量之间的相关性建立模型,如同在选择模型中那样(如 Heckman,1979)。我们将在 4.8.2 节对此进行详细讨论。

而且,有几种估计方法是建立在潜响应方程的基础之上(参见第 6 章)。这包括由穆森(Muthén,1984;Muthén and Satorra,1995)提出的受限信息方法,这在 6.7 节进行讨论;以及由肖恩伯格(Schoenberg,1985)阐述的期望最大化算法(EM algorithms)。有些估计方法模拟了潜响应。实例包括:6.4.1 节讨论的蒙特卡洛期望最大化算法(Monte Carlo EM algorithms)(Meng and Schilling,1996)、6.11.5 节举例的马尔可夫链蒙特卡洛方法(Markov Chain Monte Carlo methods)(Albert and Chlib,1993)、模拟动差估计法(method of simulated moments)(McFadden,1989)以及 6.3.3 节讨论的 GHK 模拟最大似然法(GHK method for simulated maximum likelihood)(Train,2003)。

此外,我们将在第 5 章的分析中看到,运用潜响应方程来探究模型辨识更为容易。最后,建立在潜响应基础上的模型诊断程序,已经由切舍尔和艾里什(Chesher and Irish,1987)以及古力洛克斯等人(Gourieroux et al.,1987a)在频率论的框架中进行了发展,以及由阿尔伯特和奇布(Albert and Chib,1995)在贝叶斯模型中进行了发展(参见 8.6 节)。

## 1.7 产生灵活分布

潜变量对于产生具有期望的方差函数与形式的分布或者具有特定依存结构的多元分布十分有用。

正如在 1.4 节所讨论的,事件数的过度离散可以通过在模型中纳入一个随机截距而得到处理(Breslow and Clayton,1993)。如果存在数值"0"过多,可以运用零堆积泊松

(zero-inflated Poisson,ZIP)模型,它是一个泊松模型与零的质量的混合(Lambert,1992)。如果响应类似于一周内的酒类消耗量,这种类型的模型是有用的。在这种情况中,一个零响应可能是由于某个人是一名"禁酒主义者"(不喝酒的人)或者仅仅因为一名饮酒者的随机波动。在ZIP模型中,构成的"隶属关系"标签("membership" label)(或者潜类别)可以被视为一个离散潜变量的实现结果。11.2节给出了这一模型关于坏牙、缺牙或补牙数量的一个应用。很多其他类型的混合模型已经应用于产生灵活分布,包括正态分布混合模型与泊松分布混合模型,可以参考的例子有Everitt and Hand(1981)、Böhning(2000)、McLachlan and Peel(2000)。

潜变量的另一项应用是纳入响应之间相关性的一种简约的方式。一个典型的例子是3.6节所讨论的纵贯性数据分析。此外,潜变量可以用来归纳非正态响应——这里灵活多元分布是不存在的——之间的相关性。如,库尔和阿奎斯提(Coull and Agresti,2000)在对数概率回归模型的线性估计量中运用相关的随机截距来建立二项响应之间相关性的模型。他们称这些模型为二项对数概率正态模型(binomial logit-normal models,BLM)。潜变量还常常用来归纳不同过程之间的依存性。一个普通的例子是,在所谓的"共享随机效应"模型中,对兴趣的一个响应与一个缺失的数据过程构建一个联合模型(Wu and Carroll,1988)。这一思想运用于14.5节的内生试验模型的建构(endogenous treatment modeling)以及14.6节的重复测量与生存的联合模型中。

## 1.8 组合不同来源的个体单元信息

为潜变量赋值通常是很有意义的,即对连续的潜变量以及类别事件中的类别和种类采取记分的形式。这种预测、得分或分类对于讨论过的所有类型的潜变量都是重要的。例如,类别变量真实分类的包含误差的测量(或错误分类)在医学诊断中是至关重要的,这在9.3节关于心肌梗塞的例子中有所展示。9.4节阐述了假设构念的得分,比如能力在教育中是主要的,然而特定单元效应的预测是小域估计的目标,11.4节对疾病制图进行讨论。

将得分与分类建立在明确的潜变量模型之上是非常有益的。这一方法允许对测量的信度与效度进行经验性的评估,并且提供组合信息的最佳方法。正如在第7章所解释的,针对一个单元的预测值不是单独地基于对那一单元的测量,而且受到针对单元总体的潜变量估计分布的效应。在经验性的贝叶斯方法中,潜变量分布代表着一种(经验的)优先,而给定潜变量的情况下,测量的条件分布则代表着"似然"。因为潜变量分布是通过运用来自所有单元的信息进行估计的,所以对一个给定单元的预测就可以从其他单元的测量中"借助力量"(Rubin,1983;Morris,1983)。例如,在接近地理单元的疾病制图中,提供了有用的信息,可以提高基于少数事件而进行的对发病率的估计。

## 1.9 总　结

我们已经论证,潜变量遍及现代统计学,并且描述了它们是如何用来阐述广泛不同的现象,如真实变量含误差的测量、假设构念、未观测到的异质性、缺失值、反事实以及类别变量背后的潜响应。此外,潜变量可以用于产生灵活的多元分布,并且可以统合来源不同的个体单元信息。

在下一章中,我们描述广义线性模型的种类,它们尽管并未包含潜变量,但却为本书其他章节的核心内容——潜变量模型——提供了必要的基础。

# 2 对不同响应过程建模

## 2.1 介 绍

本章我们将介绍很多种响应过程，这些观测响应的类型如下：

- 连续响应或定距响应。
- 二项响应。
- 分组响应。
- 删失(censored)响应。
- 定序响应。
- 无序多项(polytomous)响应或多类别响应。
- 成对比较(pairwise comparisons)。
- 排序(rankings)或排列(permutations)。
- 事件数(counts)。
- 存续或生存。

统计建模的目的是为了抓住所研究的经验过程的主要特征(进一步的深入讨论请见8.2节)。最典型的是，第一个简单步骤聚焦于一组有限的响应变量，并通过给定一组解释变量来考察这些变量的数据产生过程。单变量模型只有一个响应变量，而多变量模型则有几个响应变量，甚至可能包括同时充当响应变量和解释变量的中介变量或中间变量。响应变量有时被称为“因变量”“内生变量”或“结果变量”，而解释变量则被称为“自变量”“外生变量”或“预测变量”。首要关注的解释变量有时被称为“显露(exposures)”或“(风险)因子”，而其他的则被称为“混淆因子(confounders)”或“协变量”。但是，“协变量”这一术语通常被用来作为解释变量的一个一般性术语。

变量可以根据它们的“测量层次”进行进一步的分类。对于解释变量，区分连续变量或类别变量通常是容易实现的，因为我们并不对它们建立模型，而只是以它们为条件。如果变量的取值是有序的，并且数值之间的差异是有意义的，这一变量可以被视为典型的连续变量，否则就是类别变量。

对于响应变量，我们还必须考虑可能产生这一响应的过程，因为这对于设计一个适当的统计模型来说是很关键的。如，一个变量呈现出有序的离散值，如1,2,3等，可以表示一个诸如疼痛程度(没有、轻微、中度等)的定序响应，也可以表示一周中头痛间歇期天数的这样一个事件数，或者表示从一些诊断得出的接近死亡的月数的这样一个离散时间

的存续(discrete-time duration)。很明显,产生这些相同变量值的**响应过程**具有完全不同的性质,因而需要建立不同的统计模型。

有两种一般性的方法可以用来建立响应过程的模型。在统计学和生物统计学中,最普遍的方法是广义线性模型,但是在计量经济学和心理测验学中,比较流行的是潜响应方程。尽管它们的表现形式截然不同,但是这些方法可以产生针对很多响应类型的等价模型。然而,正如我们将要在后续章节中所看到的,方程式的选择将会涉及估计与辨识问题。潜响应方程是非常有用的,即使是针对这样一些应用:其中,依据一个潜响应所做出的解释表现为人为设计的。

我们首先介绍广义线性模型及其扩展。然后我们介绍潜响应方程,并指出这些方法之间的一致之处。最后,将对存续或生存数据进行单独的讨论,因为它们并不完全适合这两种架构中的任何一个。连续时间模型与离散时间模型都在考虑之列。

在本章中,我们仍然没有介绍随机系数或公共因子。然而,本章讨论的模型将为第4章所介绍的一般模型架构提供基本的素材。

## 2.2 广义线性模型

### 2.2.1 介 绍

在广义线性模型(如 Nelder and Wedderburn,1972)中,解释变量仅仅通过单元 $i$ 的**线性估计量** $\nu_i$ 对响应产生作用:

$$\nu_i = \mathbf{x}_i'\beta$$

其中,$\mathbf{x}_i$ 是解释变量的一个向量,$\beta$ 包含相应的回归参数。

连续型解释变量和类别解释变量都能被纳入到模型当中。对于年龄之类的连续变量,通常用一个单项来表示年龄对于线性估计量的线性效应。对这种效应进行建模的其他方式还包括多项式、样条线(splines)或其他平滑函数(smooth functions)。对于民族之类的类别变量(如 Norwegian,German,other),除设定的一个基准类之外,每一类别都要明确地设置一个虚拟变量,比如"其他"。例如,关于挪威人的一个**虚拟变量**,如果此人是挪威人,变量值等于1,否则为0,因此,相应的系数代表着是挪威人与"其他"民族相比较时的效应。

通过设定给定的线性估计量 $y_i$ 的条件概率(密度),响应过程就能得到充分的描述。最简单的响应过程是连续响应的过程。在这种情形下通常设定**线性回归模型**

$$y_i = \nu_i + \epsilon_i \tag{2.1}$$

其中,残差 $\epsilon_i$ 服从均值为零且固定方差(constant variance)为 $\sigma^2$ 的独立正态分布:

$$\epsilon_i \sim N(0,\sigma^2) \tag{2.2}$$

线性回归模型的特殊情况包括方差分析(ANOVA)与协方差分析(ANCOVA)模型。

线性回归模型也可以通过设置响应的条件期望(conditional expectation)来进行定义,给定线性估计量 $\nu_i$,就等于 $\nu_i$:

$$\mu_i \equiv E(y_i \mid \nu_i) = \nu_i$$

其中,$y_i$ 服从均值为 $\mu_i$、方差为 $\sigma^2$ 的独立正态分布。

对于采用0或1值的二项响应或二元响应(binary responses),响应为1的条件概率 $\Pr(y_i = 1 \mid \nu_i)$,正好是 $y_i$ 的条件期望 $\mu_i$。对此可以建立一个**对数概率回归**:

$$\mu_i = \frac{\exp(\nu_i)}{1+\exp(\nu_i)} \quad \text{或者} \quad \ln\frac{\mu_i}{1-\mu_i} = \nu_i$$

或者一个**正态概率回归**

$$\mu_i = \Phi(\nu_i) \quad 或者 \quad \Phi^{-1}(\mu_i) = \nu_i$$

其中，$\Phi(\cdot)$是标准正态累积分布函数。以$\nu_i$为条件，$y_i$服从独立伯努利分布。

事件数是赋予响应的离散非负整数(0,1,…)。事件数的标准模型是包含期望值的**泊松回归**(Poisson regression)：

$$\mu_i = \exp(\nu_i) \quad 或者 \quad \ln(\mu_i) = \nu_i$$

以及泊松分布

$$\Pr(y_i \mid \mu_i) = \frac{\exp(-\mu_i)\mu_i^{y_i}}{y_i!} \tag{2.3}$$

如果事件是针对一个单元进行计算的，这些事件在一段连续的时间内以一个固定比率发生，并且相互独立，那么事件数就属于一个泊松分布。

如果一个事件数相当于在一个给定的$n$次"试验"中所发生的事件数量(或一个事件的多种机会)，并且针对一个单元的这些事件是独立的、具有同等可能性的，该事件数就属于一个二项分布。$n$个当中的一个比率$y_i$的概率，即

$$\Pr(y_i \mid \mu_i) = \binom{n}{y_i n} \mu^{y_i n}(1-\mu_i)^{(1-y_i)n}$$

其中，二项系数$\binom{n_i}{y_i n} = n!/[(y_i n)!\ (n-y_i n)!]$是在$n$个对象中选择$y_i n$的方法数，不考虑它们的次序关系。

### 2.2.2 模型结构

上文谈到的所有模型都具有一个共同的结构并且都代表广义线性模型的特殊情形，广义线性模型由两个要素进行界定：

(1)响应的期望值与线性估计量之间的函数关系为

$$\mu_i = g^{-1}(\nu_i) \quad 或者 \quad g(\mu_i) = \nu_i$$

其中，$g(\cdot)$是一个**链接函数**(link function)。我们已经分别遇到关于线性回归、对数概率回归、正态概率回归和泊松回归的同一性链接(identity link)、对数概率链接、正态概率链接和对数链接(log link)。这些链接及其他常见的链接如表2.1所示。

**表2.1 常见的链接函数**

| 链接 | $g(\mu)$ | $g^{-1}(\nu)$ | $g^{-1}(\nu)$的范围 |
|---|---|---|---|
| 同一性链接(Identity) | $\mu$ | $\nu$ | $-\infty,\infty$ |
| 双向链接(Reciprocal) | $\frac{1}{\mu}$ | $\frac{1}{\nu}$ | $-\infty,\infty$ |
| 对数链接(Logarithm) | $\ln(\mu)$ | $\exp(\nu)$ | $0,\infty$ |
| 对数概率链接(Logit) | $\ln\left(\frac{\mu}{1-\mu}\right)$ | $\frac{\exp(\nu)}{1+\exp(\nu)}$ | 0,1 |
| 正态概率链接(Probit) | $\Phi^{-1}(\mu)$ | $\Phi(\nu)$ | 0,1 |
| 尺度正态概率链接(Scaled probit) | $\sigma\Phi^{-1}(\mu)$ | $\Phi\left(\frac{\nu}{\sigma}\right)$ | 0,1 |
| 余重对数链接(Complementary log-log) | $\ln(-\ln(1-\mu))$ | $1-\exp(-\exp(\nu))$ | 0,1 |

(2)响应的条件概率分布,是期望值为 $\mu_i$ 并且可能具有一个公共尺度参数 $\phi$(common scale parameter)的指数族的一个部分,即

$$f(y_i \mid \theta_i,\phi) = \exp\left\{\frac{y_i\theta_i - b(\theta_i)}{\phi} + c(y_i,\phi)\right\} \tag{2.4}$$

其中,$\theta_i$ 是标准参数(canonical parameter)或自然参数(natural parameter),$\phi$ 是尺度参数(scale parameter)或离散参数(dispersion parameter),$b(\cdot)$和 $c(\cdot)$是取决于指数族中某种分布的函数。我们已经遇到正态或高斯分布、伯努利分布、泊松分布与二项分布。表2.2对这些分布以及指数族中的其他重要分布进行了详细说明。

假定各单元的响应是独立的,则广义线性模型的似然为

$$l = \prod_{i=1}^{N} l_i$$

其中,$l_i \equiv f(y_i \mid \theta_i,\phi)$是第 $i$ 个单元($i=1,2,\cdots,N$)对似然的贡献,对数似然则为

$$\ell = \sum_{i=1}^{N} \ell_i$$

其中,$\ell_i \equiv \ln l_i$。关于 $\theta_i$ 的对数似然贡献的一阶和二阶导数分别为

$$\frac{\partial \ell_i}{\partial \theta_i} = \frac{[y_i - b'(\theta_i)]}{\phi} \tag{2.5}$$

且

$$\frac{\partial^2 \ell_i}{\partial \theta_i^2} = \frac{-b''(\theta_i)}{\phi} \tag{2.6}$$

其中,$b'(\theta_i)$和 $b''(\theta_i)$是 $b(\cdot)$在 $\theta_i$ 处的一阶和二阶导数。在6.8.1节,我们将介绍运用迭代重加权最小二乘(iteratively reweighted least squares)的广义线性模型极大似然估计法。

### 2.2.3 均值函数与链接函数的选择

根据标准似然理论,期望得分为0,即

$$\mathrm{E}\left(\frac{\partial \ell}{\partial \theta_i}\right) = 0$$

故它服从方程式(2.5),即

$$\mu_i = b'(\theta_i)$$

将 $\theta_i$ 写成 $\mu_i$ 的一个函数,得到**标准链接**函数为

$$g(\mu_i) = \theta_i(\mu_i)$$

标准链接具有非常便利的统计特性。然而,对于链接的选择必须接受理论思考与模型拟合的指导。选择一个链接函数的一种考虑是取值的范围,当 $-\infty \leqslant \nu_i \leqslant \infty$ 时,均值 $\mu_i = g^{-1}(\nu_i)$(参阅表2.1)。如,对数概率和正态概率链接对于二项响应来说是非常流行的,因为它们将概率 $\mu_i$ 严格限定在允许的区间[0,1]之内。相反,在这种情况下运用同一性链接,则可能导致负的或过大的预测性概率。

另一个重要的考虑与回归参数的解释有关。因为 $\nu_i = \mathbf{x}_i'\beta$,所以可以用一个同一性链接来表示协变量在均值上的叠加效应(additive effects),$\mu_i = \mathbf{x}_i'\beta$;用一个对数链接表示乘积效应(multiplicative effects),$\mu_i = \exp(\mathbf{x}_i'\beta)$。对二项响应可以采用一个对数概率链接,得到一个关于比率(odds)的乘积模型,$\mu_i/(1-\mu_i) = \exp(\mathbf{x}_i'\beta)$。这一链接在病例对照研究(case-control studies)中非常有用,因为优比(odds-ratios)对于回溯性抽样(retrospective sampling)或"择基"抽样("choice-based"sampling)来说是守恒的(Farewell,1979)。

表 2.2　指数族的构成项

| 分　布 | 标准链接 $\theta(\mu)$ | 累积函数 $b(\theta)$ | 离散参数 $\phi$ | 期望值 $b'(\theta)$ | 方差 $\phi b''(\theta)$ | 概率或密度 $f(y\|\theta,\phi)$ |
|---|---|---|---|---|---|---|
| 伯努利分布(Bernoulli) | $\ln(\mu/(1-\mu))$ | $\ln(1+\exp(\theta))$ | 1 | $\frac{\exp(\theta)}{1+\exp(\theta)}$ | $\mu(1-\mu)$ | $\mu$ |
| 二项分布⁺(Binomial⁺) | $\ln(\mu/(1-\mu))$ | $\ln(1+\exp(\theta))$ | $1/n$ | $\frac{\exp(\theta)}{1+\exp(\theta)}$ | $\mu(1-\mu)/n$ | $\binom{n}{yn}\mu^{yn}(1-\mu)^{(1-y)n}$ |
| 泊松分布(Poisson) | $\ln(\mu)$ | $\exp(\theta)$ | 1 | $\exp(\theta)$ | $\mu$ | $\frac{\mu^y\exp(-\mu)}{y!}$ |
| 正态分布(Normal) | $\mu$ | $\theta^2/2$ | $\sigma^2$ | $\theta$ | $\sigma^2$ | $\frac{1}{\sqrt{2\pi\sigma^2}}\exp\frac{-(y-\mu)^2}{2\sigma^2}$ |
| 伽玛分布(Gamma) | $-1/\mu$ | $-\ln(-\theta)$ | $\alpha^{-1}$ | $-1/\theta$ | $\mu^2\alpha^{-1}$ | $\frac{1}{\Gamma(\alpha)}\left(\frac{\alpha}{\mu}\right)^{\alpha}y^{\alpha-1}\exp\left(-\frac{\alpha y}{\mu}\right)$ |
| 逆高斯分布(Inverse Gaussian) | $1/\mu^2$ | $-(-2\theta)^{1/2}$ | $\sigma^2$ | $(-2\theta)^{-1/2}$ | $\mu^3\sigma^2$ | $\frac{1}{\sqrt{2\pi y^3\sigma^2}}\exp\left\{\frac{-(y-\mu)^2}{2(\mu\sigma)^2 y}\right\}$ |

注：⁺$y$ 为 $n$ 个“试验”中“成功”的比例。

在流行病学中，受因果关系这一特定概念的推动，针对二项响应采用同一性链接得到了提倡（参阅 Skrondal（2003）及其中的参考书）。这说明在有些时候，会有一些原因导致偏离指数族的标准链接。

### 2.2.4 方差函数与分布的选择

分布的选择取决于响应变量的类型、可能产生响应变量的过程以及经验分布的形态。对于二项响应，显而易见的选择是伯努利分布。就事件数而言，如果产生事件的过程具有确定的特征（固定的发生比（constant incidence rate）与独立性），则可以看出它具有一个泊松分布。

分布的选择也决定了响应的条件方差是均值一个函数。它是根据标准似然理论得出的，即

$$-\mathrm{E}\left(\frac{\partial^2 \ell}{\partial \theta_i^2}\right) = \mathrm{E}\left(\frac{\partial \ell}{\partial \theta_i}\frac{\partial \ell}{\partial \theta_i}\right)$$

用方程式（2.5）和方程式（2.6）的术语替换后，得

$$\mathrm{Var}(y_i \mid \nu_i) = \phi b''(\theta_i) = \phi V(\mu_i)$$

其中，$V(\mu_i)$就是方差函数，$\phi$ 是离散参数。如，对于泊松分布，这个方差等于均值；而对于正态分布来说，方差是由一个固定参数 $\phi = \sigma^2$ 给出的（参阅表 2.2）。

## 2.3 广义线性模型的扩展

### 2.3.1 过少离散（underdispersion）与过度离散的建模

事件数典型地通过二项分布或泊松分布进行建模（在事件仅仅以预先确定的“试验”次数出现时是二项分布，$n$）。对于这两种分布，在给定解释变量的情况下，条件方差是由均值决定的。然而，实际观测的条件方差经常大于或小于模型的计算值，这就是我们所知道的两种现象，分别是过度离散与过少离散。过度离散可能是因为二项概率或泊松比（Poisson rates）的变异性没有被纳入的协变量所充分解释，并且过度离散比过少离散更为普遍。

过度离散与过少离散的一种特别的解决方法是为方差引入一项额外的比例参数 $\phi^+$，得出针对二项分布：

$$\mathrm{Var}(y_i \mid \nu_i) = \phi^* \mu_i(1 - \mu_i)/n$$

以及针对泊松分布：

$$\mathrm{Var}(y_i \mid \nu_i) = \phi^* \mu_i$$

注意，这些设定并不需要与概率模型相一致，但仍然可以运用所谓的**准似然**方法（quasi-likelihood methods）进行估计（参阅 6.8 节）。

另一种针对过度离散的“适当的”建模方法是，允许均值在各个单元之间以固定的协变量值随机地变动。结合二项响应分布与关于概率的贝塔分布（beta distribution），可以得到贝塔-二项分布（beta-binomial distribution）。负二项分布（negative binomial distribution）是将泊松响应分布与关于比率的伽玛分布（gamma distribution）相结合而得到的。另一种可能性是在线性估计量中纳入一个正态分布的随机截距，广义模型架构的一个专门案例将在第 4 章进行介绍。林赛（Lindsey，1999：197-220）在其著作中对过度离散与过少离散建模的这些方法以及一些更深入的方法进行了讨论。

### 2.3.2 异方差性建模(Modeling heteroscedasticity)

线性回归模型中的一个传统假设是异方差性(heteroscedasticity),也就是说残标准差(residual standard deviation)$\sigma$ 被假设为在各个单元上是固定的。然而,$\sigma$ 也可能取决于分类协变量或连续协变量。如,当比较 11 岁的男孩与女孩的身高时,我们会预期女孩的身高变动性更大,因为很多女孩可能已经进入青春期,而所有男孩可能都还处于青春期之前。由于标准差必须是正的,因此,运用一个对数链接对异方差性进行建模非常便利,即

$$\ln \sigma_i = \mathbf{x}_i'\boldsymbol{\iota} \quad \text{或} \quad \sigma_i = \exp(\mathbf{x}_i'\boldsymbol{\iota}) \tag{2.7}$$

其中,$\mathbf{x}_i$ 是协变量,$\boldsymbol{\iota}$ 是参数。这种乘积异方差性(multiplicative heteroscedasticity)是由哈维(Harvey,1976)提出的。这一设定还可以运用于含有尺度(scale)或离散参数的其他模型;同时,可参阅 2.3.4 节中的内容。

### 2.3.3 多项响应模型

多项响应是无序的分类响应,如一次选举中的政党投票。在计量经济学中,这类响应经常被认为是**离散选择**(discrete choices)。在统计学中,应用的术语则包括定性响应或定类响应,因为这些类别不具有数量关系,且没有内在次序。其他的术语还包括量子(quantal)响应、多项(polychotomous)响应或多类别(multinomial)响应。

针对每一个类别 $a_s, s=1,\cdots,S$,设定一个单独的线性估计量。在这种情况下,响应可以被视为多元的,并且通常用一个 01 向量来表示,其中 1 表示实现的类别(realized category)、0 表示其他类别(Fahrmeir and Tutz,2001)。某一事物的第 $s$ 个类别或备选项 $a_s$ 的概率,通常可以建立一个**多项对数概率**模型:

$$\Pr(y_i = a_s) = \frac{\exp(\nu_i^s)}{\sum_{t=1}^{S}\exp(\nu_i^t)} \tag{2.8}$$

其中,$\nu_i^s$ 是单元 $i$ 和类别 $a_s$ 的线性估计量,方程式下面的和包括了所有的 $S$ 个类别。蒂森和斯滕伯格(Thissen and Steinberg,1986)将这种类型的模型称为**除总模型**(divide-by-total model)。注意 $\Pr(y_i = a_s)$ 是给定线性估计量的一个条件概率,尽管我们在这里已经抑制了条件作用(conditioning),并且为了简单起见在本章的剩余部分都做同样处理。

我们可以将单元和特定分类协变量或属性(attributes)纳入线性估计量。如,考虑这样一个案例,其中,响应类别是超市,顾客的响应代表着他对超市的选择 $a_s$。该线性估计量可以纳入顾客的特定变量 $\mathbf{x}_i$,如收入;以及顾客与超市的特定变量 $\mathbf{x}_i^s$,如顾客前往超市的交通时间。这个线性估计量则成为

$$\nu_i^s = m^s + \mathbf{x}_i'\beta^s + \mathbf{x}_i^{s\prime}\beta \tag{2.9}$$

其中,$m^s$ 是一个特定类别的常数(constant),$\beta^s$ 是特定单元协变量 $\mathbf{x}_i$ 的特定类别效应,而 $\beta$ 是单元和特定类别协变量 $\mathbf{x}_i^s$ 的固定效应(constant effects)。如果类似这种情况,即交通时间的效应对于小型超市的效应大于大型超市,那么 $\mathbf{x}_i^s$ 的系数在类别之间也可能是不同的。

应当注意的是,在单元 $i$ 的线性估计量中加入一个 $B_i$ 项,方程式(2.8)中的 $\nu_i^s(s=1,\cdots,S)$ 并不会改变其概率,因为它实际上是分子与分母同时乘以 $\exp(B_i)$。由此,我们可以将常数加进 $m^s$ 和 $\boldsymbol{\beta}^s$,而不需要改变模型。这是模型**辨识问题**的一个实例(同时参阅第 5 章)。该问题可以通过将一种类别作为"基准类(base category)"(一般是第一个,$a_1$)

并且强制 $m^1=0,\boldsymbol{\beta}^1=\mathbf{0}$ 而得到解决。

计量经济学中标准的**条件对数概率模型**,是作为一种特殊的情形出现,其中没有特定单元协变量 $\mathbf{x}_i$,也没有常数项 $m^s$(请参见 McFadden,1973)。在诸如生物统计学中的一种标准模型——**多项对数概率回归模型**(请参见 Hosmer and Lemeshow,2000),属于这种特殊情形,其中没有特定类别协变量 $\mathbf{x}_i^s$。

### 2.3.4 定序响应模型

**累积模型**

有序分类响应或定序响应的模型可以通过将累积概率 $\Pr(y_i \leqslant a_s)$ 链接到线性估计量而进行界定(McCullagh,1980):

$$g[\Pr(y_i \leqslant a_s)] = \kappa_s - \nu_i \qquad s = 1,\cdots,S-1 \tag{2.10}$$

其中,$a_1<a_2<\cdots<a_s$ 是有序的响应类别,$\Pr(y_i \leqslant a_s)=1$ 且 $\kappa_s$ 为**阈值参数**(threshold parameters),$\kappa_1<\kappa_2<\cdots<\kappa_s-1$。最有代表性的链接函数包括正态概率、对数概率和余重对数。这类模型通常称为累积模型(cumulative model)或等级响应模型(graded response model)(Samejima,1969)。

考虑方程式(2.10)右侧的 $\kappa_s-\nu_l$。因为添加一个任意的常数 $\beta_0$ 到线性估计量,可以通过为每个 $\kappa_s$ 添加相同的常数而得到平衡,所以很明确的是,我们不能同时估计常数与所有的阈值。这一辨识问题可以通过设定 $\kappa_1=0$ 而加以解决,使这一参数化等同于前面章节中的二项响应模型所运用的那样(假设 $S=2,a_1=0$ 且 $a_2=1$)。或者,如果我们改为从线性估计量中略去常数项,则 $\kappa_1$ 可以是一个模型参数。

根据方程式(2.10)可知,一个特定响应 $y_s$ 的概率是

$$\Pr(y_i = a_s) = \Pr(y_i \leqslant a_s) - \Pr(y_i \leqslant a_{s-1}) \tag{2.11}$$

促使蒂森和斯滕伯格(Thissen and Steinberg,1986)将这些模型称为**差分模型**(difference model)。

在方程式(2.10)中,累积响应概率上的协变量效应在类别 $s$ 之间是恒定的,表现为一种被称为平行回归假设(parallel regression assumption)的特性。对于一个对数概率链接,$y$ 大于 $a_s$ 的比率为

$$\frac{\Pr(y_i > a_s)}{\Pr(y_i \leqslant a_s)} = \frac{1-\Pr(y_i \leqslant a_s)}{\Pr(y_i \leqslant a_s)} = \exp(\boldsymbol{x}_i'\boldsymbol{\beta} - \kappa_s)$$

对于两个单元 $i$ 和 $i'$,其比率的比值 $\exp[(\mathbf{x}_i-\mathbf{x}_{i'})'\boldsymbol{\beta}]$,对于所有类别 $s$ 来说都是相同的,即一种称为**比例比率**(proportional odds)的特性。累积模型的一个有用的特征是,估计的回归参数在多种类别合并后都是近似恒定不变的。

响应类别之间的协变量固定效应假设,可以通过让阈值取决于协变量 $\mathbf{x}_{2i}$ 而得到释放(Terza,1985),即

$$\kappa_{is} = \mathbf{x}_{2i}'\boldsymbol{\zeta}_s$$

其中,$\boldsymbol{\zeta}_s$ 是一个参数向量。考虑正态概率形式,这一模型则变为

$$\mathrm{P}(y_i \leqslant a_s) = \Phi(\kappa_{si}-\nu_i) = \Phi(\mathbf{x}_{2i}'\boldsymbol{\zeta}_s - \mathbf{x}_i'\beta) \tag{2.12}$$

其中,$\mathbf{x}_{2i}$是有特定类别效应的协变量,$\mathbf{x}_i$ 是有固定效应的协变量。很明显,$\mathbf{x}_{2i}$和 $\mathbf{x}_i$ 中包含的任何变量的系数都不是被单独地辨识出来的。这种参数化存在一个问题,即对于某些协变量值,阈值并不必然满足这个顺序约束(order constraint)$\kappa_{i1}<\kappa_{i2}<\cdots<\kappa_{is-1}$,因此方程式(2.11)中的概率并未被限定为非负数。顺序约束可以通过以下参数化方法加以实现,即

$$\kappa_{i1} = 0, \kappa_{is} = \kappa_{is-1} + \exp(\mathbf{x}_{2i}'\boldsymbol{\zeta}_s) \qquad s = 2, \cdots, S$$

可参阅 Fahrmeir and Tutz(2001)。

释放平行回归假设的另一种可选择的策略是采用一种**尺度定序正态概率链接**(scaled ordinal probit link),其中,尺度参数如方程式(2.7)中那样进行建模:

$$\ln \sigma_i = \mathbf{x}_{2i}'\boldsymbol{\iota}$$

其中,$\boldsymbol{\iota}$ 是一个参数向量。然后模型可以写为

$$P(y_i \leqslant a_s) = \Phi[(\kappa_s - \nu_i)/\sigma_i] = \Phi(\kappa_{si}^* - \nu_i^*) \tag{2.13}$$

其中,$\kappa_{si}^* = \kappa_s/\sigma_i, \nu_i^* = \nu_i/\sigma_i$。类似这种异方差正态概率模型在格里尼(Greene,2003:680)关于二项响应的讨论中有所论述,并且斯科隆多(Skrondal,1996)将其应用于定序响应。尺度阈值(scaled thresholds)具有下述形式:

$$\kappa_{si}^* = \kappa_s \exp(-\mathbf{x}_{2i}'\boldsymbol{\iota})$$

而尺度线性估计量的形式为

$$\nu_i^* = \mathbf{x}_i'\boldsymbol{\beta}\exp(-\mathbf{x}_{2i}'\boldsymbol{\iota})$$

为了辨识,模型中关于 $\sigma$ 的截距是省略的,因为向 $\mathbf{x}_{2i}'\boldsymbol{\iota}$ 添加一个常数 $\boldsymbol{\iota}_0$ 可以通过其他方法进行抵消,即用 $\exp(-\boldsymbol{\iota}_0)$ 来除 $\kappa_s$ 和 $\boldsymbol{\beta}$。如果 $\mathbf{x}_i$ 与 $\mathbf{x}_{2i}$ 之间有交集,该模型可能会由于为 $\mathbf{x}_i$ 与 $\mathbf{x}_{2i}$ 指定不同的函数形式而得到辨识,但这种辨识很可能是脆弱的。

**铅版模型与邻类对数概率模型**(Stereotype and adjacent category logit.model)

除了建立累积概率模型之外,我们还可以运用方程式(2.8)与方程式(2.9)中的多项对数概率模型——包含反映类别顺序的额外参数约束——来设定概率 $\Pr(y_i = a_s)$ 的模型。

**铅版模型**(Anderson,1984)是通过强加约束 $\boldsymbol{\beta}^s = \alpha^s\boldsymbol{\beta}$ 和 $\alpha^1 < \cdots < \alpha^s$ 而获得的,因而其线性估计量为

$$\nu_i^s = m_s + \alpha^s(\mathbf{x}_i'\boldsymbol{\beta})$$

除常用的约束 $m^1 = 0$ 和 $\alpha^1 = 0$ 之外,诸如 $\alpha^2 = 1$ 的进一步约束对于设定 $\alpha^s$ 的尺度是必要的,因为只有 $\alpha^s\boldsymbol{\beta}$ 乘积的尺度是可辨识的。注意,铅版模型并不包含特定类别的协变量。

在铅版模型中,设 $\alpha^s = s$,得到邻类对数概率模型。在该模型中,相邻类别的比率为

$$\frac{\Pr(y_i = a_s)}{\Pr(y_i = a_{s-1})} = \exp(m^s - m^{s-1} + \boldsymbol{x}_i'\boldsymbol{\beta})$$

因此,对于两个个体 $i$ 与 $i'$,相邻类别的优比为

$$\exp[(\mathbf{x}_i - \mathbf{x}_{i'})'\boldsymbol{\beta}]$$

这对于所有 $s$ 个类别都是相同的。注意的是,邻类对数概率模型估计的是相邻类别比率的概率,而比例比率模型估计的是累积比率的概率。邻类对数概率的线性估计量

$$\nu_i^s = m^s + s\mathbf{x}_i'\boldsymbol{\beta}$$

与标准多项多数概率模型的线性估计量——含有特定类别协变量 $\mathbf{x}_i^s = s\mathbf{x}_i$——具有相同的形式。古德曼(Goodman,1983)将邻类对数概率模型称为平行比率模型(parallel odds model),而在题器响应理论中,该模型是为大家所熟知的部分评分模型(partial credit model)(Masters,1982)(参阅 3.3.4 节)。

代替向 $\alpha^s$ 赋予等距的分值,我们可以指定反映有序的类别之间"距离"的分值。铅版模型可以被认为是邻类对数概率模型的一般化,当中的分值需要进行估计而不是固定的。在这两个模型中,如果两个 $\alpha^s$ 和一个有序的解释变量都被指定了分值,那么它们就相当于线性关联模型(linear by linear association model)(Goodman,1979;Agresti,2002)。

**升级比对数概率模型**(Continuation ration logit model)

另一种可能是在给定 $y_i$ 大于等于 $a_s(y_i \geqslant a_s)$ 的情况下,估计大于类别 $a_s$ 的比率的概率为

$$\frac{\Pr(y_i > a_s)}{\Pr(y_i = a_s)} = \exp(m^s + \mathbf{x}_i'\boldsymbol{\beta})$$

即为"阶段(stage)" $a_s$ 以上的各个连续级别的比率与截止到该阶段的比值。一个等价模型(equivalent model)并非是通过对类别的次序进行反转而获得的,这意味着升级比模型只能应用于连续的阶段。由于它常常应用于离散时间存续(连续的阶段)的分析,因此,我们将在2.5.2节继续讨论这一模型。

### 2.3.5 复合链接(Composite links)

复合链接的形式为

$$\mu_i = \sum_k c_{ik} g_k^{-1}(\nu_{ik}) \tag{2.14}$$

其中,$c_{ik}$是已知的常数。这种链接函数对于事件数数据的建模是十分重要的,其中某些观测事件数代表着不同协变量值的事件数总和,典型的是由于协变量信息缺失。

一个著名的例子是血型问题,即如果母亲与父亲贡献基因 A 和 O,在任何可能的组合 AA,AO 或 OA 中,后代遗传血型(表现型)为 A 的概率问题。如果 A 和 O 的基因频数是 $p$ 和 $r$,则血型 A 的期望频数为

$$N(p^2 + 2pr) = \exp(\ln N + 2\ln p) + \exp(\ln N + \ln p + \ln r) + \exp(\ln N + \ln p + \ln r)$$

其中,等号右侧的3个关系式等于基因型 AA,AO 与 OA 的期望频数。在这个例子中,方程式(2.14)中的 $i$ 表示观测的表现型,而 $k$ 表示未观测到的基因型。如果所有的基因型 $k$ 与表现型 $i$ 是一致的,则系数 $c_{ik}$为1,否则系数 $c_{ik}$为0,$g^{-1}$是指数函数,单元总数的对数 $\ln N$ 被设置为一种偏移量(offset)(系数设定为1的一个协变量),$\ln p$ 和 $\ln r$ 是模型参数。

复合链接的另一个例子是定序响应的累积模型中一种响应类别的概率,它等于累积概率之间的一个差值(difference)(参见方程式(2.11))。在9.4节中,我们将运用一个复合链接去设定一个三参数题器反应模型。关于复合链接应用的深入案例,我们可以参考:Thompson and Baker(1981)、Rindskopf(1992)。

## 2.4 潜响应方程(Latent response formulation)

一个响应变量 $y_i$ 通常可以视为一个部分观测(partial observation),或者是一个连续的**潜响应** $y_i^*$ 的粗化(coarsening)(如 Pearson,1901)。设潜响应的模型为

$$y_i^* = \nu_i + \epsilon_i$$

其中,$\nu_i$ 是一个线性估计量,$\epsilon_i$ 是一个误差项或干扰项。对于连续响应来说,潜响应简单地等同于观测响应,$y_i = y_i^*$。当潜响应通过应用以下各个小节所描述的不同类型的阈值函数进行粗化后,就出现了多种其他响应类型。

### 2.4.1 分组、区间删失(interval-censored)、定序和二项响应

观测响应 $y_i$ 任取 $S$ 个响应类型 $a_s$ 中的一种,$s=1,\cdots,S$,则观测响应与潜响应之间的关系可以表述为

$$y_i = \begin{cases} a_1 & 当 \quad \kappa_{i0} < y_i^* \leqslant \kappa_{i1} \\ a_2 & 当 \quad \kappa_{i1} < y_i^* \leqslant \kappa_{i2} \\ \vdots & \qquad \vdots \\ a_S & 当 \quad \kappa_{iS-1} < y_i^* \leqslant \kappa_{iS} \end{cases} \tag{2.15}$$

其中，$\kappa_{i0} = -\infty$，$\kappa_{iS} = \infty$。图 2.1 是 $S=3$、$\epsilon_i$ 服从正态分布的图解。

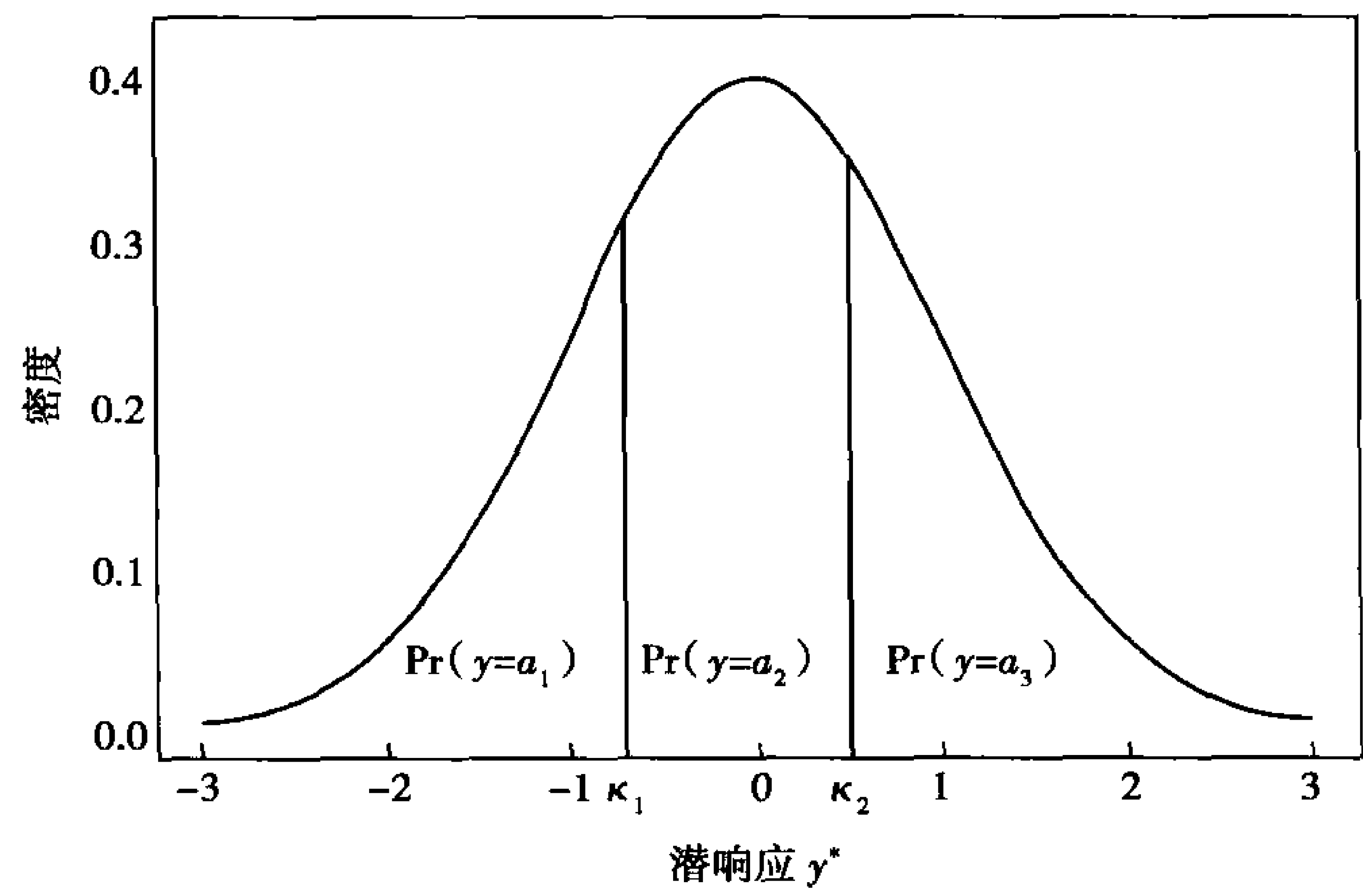



**图 2.1　分组、区间删失或定序响应的潜响应方程($S=3$)**

对于**分组响应**，阈值在单元之间是不变的，$\kappa_{is} = \kappa_s$，并且是先验性已知的(known a priori)。分组数据的一个实例是，薪水通过分界线 $\kappa_s$进行分组，进入预先设定的收入等级中，这是由斯特沃特(Stewart, 1983)提出的一种情形。

对于**区间删失响应**，$\kappa_{is}$在单元之间变动，并且是先验性已知的。如，一种疾病的发作时间，可能不会被精确地知道，而仅仅是位于两个临床患者的一个删失区间内，两个患者的时间选择在个体之间是变化的。

对于**定序响应**，阈值 $\kappa_s$是未知的参数，并且通常在单元之间是不变的。如，疼痛的强度可能被描述为“无”“中等”或“严重”。这些结果可以从字面上被看做是由疼痛强度造成的，是一个未观测到的连续潜响应，大于特定的阈值。有时我们可以释放常数阈值(constant thresholds)的假设，以便在疼痛容忍度上建立个体差别的模型。

**二项响应**通常可以被视为含有两个类别的定序响应，但有时候更适合被解释为比较响应(comparative responses)(参见 2.4.3 节)。

如果假定误差项服从一个正态分布，在两个类别的情况下，这个模型是二项正态概率模型(Bliss, 1934)，在超过两个类别时，则是定序正态概率模型(参见 2.3.4 节)。定序正态概率模型的潜响应方程是由艾勤森和西尔维(Aitchison and Silvey, 1957)提出来的，再由麦凯维和萨维那(McKelvey and Zavoina, 1975)进行了重新发掘。如果误差项被假定为具有对数累积分布函数(logistic cumulative distribution function)，则

$$F(\tau) \equiv \Pr(\epsilon_i \leqslant \tau) = \frac{\exp(\tau)}{1 + \exp(\tau)}$$

那么在两个类别的情况下，可以得到二项对数概率模型，在超过两个类别时，可以获得定序对数概率模型、比例比率模型和等级响应模型(参见 2.3.4 节)。

为了指出潜响应方程与广义线性模型方程的一致性，我们来考察二项的案例($S=2$, $a_1=0$, $a_2=1$, $\kappa_1=0$)。这一潜响应模型为

$$y_i^* = \nu_i + \epsilon_i, \quad y_i = \begin{cases} 1 & 当 \quad y_i^* > 0 \\ 0 & 其他 \end{cases}$$

它遵循：

$$\begin{aligned} E(y_i | \nu_i) &= \Pr(y_i = 1 | \nu_i) = \Pr(y_i^* > 0 | \nu_i) = \Pr(\nu_i + \epsilon_i > 0) \\ &= \Pr(-\epsilon_i \leqslant \nu_i) = F(\nu_i) \end{aligned}$$

其中，倒数第2个等式取决于 $\epsilon_i$ 密度的对称性。这里 $F = g^{-1}$ 是累积分布函数，正态概率模型中是标准正态分布，对数概率模型中是对数分布，对于余重对数链接来说则是冈贝尔分布。

## 2.4.2　删失响应

双重删失响应的阈值模型可以写为

$$y_i = \begin{cases} \kappa_{i1} & 当 \quad y_i^* \leqslant \kappa_{i1} \\ y_i^* & 当 \quad \kappa_{i1} < y_i^* \leqslant \kappa_{i2} \\ \kappa_{i2} & 当 \quad \kappa_{i2} < y_i^* \end{cases} \tag{2.16}$$

对于右删失响应的特殊情况，$\kappa_{i1} = -\infty$，对于左删失响应，$\kappa_{i2} = \infty$。

不同类型的删失响应在存续分析和生存分析中是非常突出的。右删失典型地是由于事件在观察期限结束的时候仍然没有发生。如果我们知道在观察开始之前事件已经发生，那么左删失就出现了。如果两种类型的删失都可能出现，这些响应就是双重删失响应。其他有关删失的例子还有天花板效应(ceiling effect)和地板效应(floor effect)。如，当用正确解决问题的百分比来衡量能力时，我们无法100%的对候选人进行区分。我们所能知道的，只是他们的能力超过或等于要求被达到的最大水平(天花板效应)。类似地，如果某些候选人不能解决任何问题，地板效应就出现了。对于正态潜响应(条件依赖于协变量)，托宾(Tobin，1958)介绍了单侧删失(one-sided censoring)并据此提出了**托宾模型**。一个类似的双重删失模型要归功于罗赛特和尼尔森(Rosett and Nelson，1975)，通常被表示成**双重限制正态概率模型**(two-limit probit)。

删失不应该与**截断**(truncation)相混淆。如果这样的单元——含有一个低于某个特定阈值的响应——被排除出样本，左截断就发生了。如，一项对于高血压病的临床试验，其中基线血压必须超过某个阈值，才能纳入研究当中。在存续的研究文献中，左截断是由于风险的产生与被纳入研究之间存在的延迟而造成的(因此存续很短的单元就被排除在外)，这种延迟通常被称为晚期进入或延时进入。如果仅仅是上述那些单元被纳入这样的样本中——它们的响应下降到某些阈值以下，右截断就出现了。一个计量经济学的经典案例是负所得税实验(negative income tax experiments)，其中家庭收入水平高于某一特定限制(如是贫困线的1.5倍)的家庭，被排除在收入研究之外。与删失不同，截断意味着我们没有关于那些响应超过阈值之个体的解释性变量的信息。

不同的粗化过程导致不同的变量类型，如表2.3所示。表中的各列是指粗化的类型以及阈值的类型。阈值在单元之间能够是变化的或者是恒定的，是已知的模型参数或者是未知的模型参数。变量的结果类型在行中给出。除连续变量和定序变量之外，变量类型通常被称为计量经济学中的**有限应变量**(limited dependent variables)。

表 2.3　响应类型与粗化类型

| 变量类型 | 粗化 | | | | 阈值 | |
|---|---|---|---|---|---|---|
| | $y_i^*$ | $y_i^* \leqslant \kappa_{is-1}$ | $y_i^* > \kappa_{is}$ | $\kappa_{is-1} < y_i^* \leqslant \kappa_{is}$ | 恒定/变动 | 已知/未知 |
| 连续 | √ | | | | — | — |
| 分组 | | | | √ | 恒定 | 已知 |
| 删失 | | | | √ | 变动 | 已知 |
| 定序/二项 | | | | √ | 恒定 | 未知 |
| 右删失 | √ | | √ | | 变动 | 已知 |
| 左删失 | √ | √ | | | 变动 | 已知 |
| 双重删失 | √ | √ | √ | | 变动 | 已知 |

## 2.4.3　比较响应

在 2.2.3 节中,我们考察了多项响应的模型。在实现的类别"支配"其他类别的情况下,多项响应可以被解释为比较的。

这种解释特别适合于**第一选择**(first choice)或**离散选择**(discrete choice)数据,其中决策者从一系列选项(类别)中进行选择。例如,在选举研究中,一个核心的结果变量是一名投票者的第一选择,如从一系列选项(如工党、保守党、自由党)中选择保守党。

另一种比较响应类型是**成对比较**,其中,在一个单元的每一对类别中,这些响应是支配性的类别。例如,在第一对中工党可能优先于自由党,在第二对中自由党可能优先于保守党,等等。

类别的**排列**也是一种比较响应。在决策情境中,排列可以被解释为排序,其中选项根据偏好排出了顺序。例如,政党可能就是排序的,如说自由派优于工党优于保守派。与成对比较数据相反的是,排序数据所包括的成对比较,必然是传递性的(自由派优于保守派是根据自由派优于工党和工党优于保守派得出的)。比较响应是定类的,在这个意义上,类别并不拥有一个被所有单元共享的内在顺序,如定序变量所假定的那样。我们发现,即使在没有实际涉及决策时,在比较响应中运用决策的术语也是很有用的。

比较响应的建模可以通过假定每个单元指定一个效用 $u_i^s$ 到每个选项 $a_s$ 而得到实现。效用这一术语应当广泛地被解释为选项的普及性或吸引力。对于多项响应,假定效用最大的选项被选择,即

$$y_i = a_s \quad \text{当} \quad u_i^s - u_i^t > 0 \qquad \forall t, t \neq s$$

如果我们将效用模型化为

$$u_i^s = \nu_i^s + \epsilon_i^s$$

其中,线性估计量 $\nu_i^s$ 的值对应不同选项可以是不同的,$\epsilon_i^s$ 服从独立冈贝尔(极端值 Type Ⅰ)分布

$$\Pr(\epsilon_i^s \leqslant \tau) = \exp[-\exp(-\tau)] \tag{2.17}$$

由此可以推出(McFadden,1973;Yellott,1977)一个特定选择 $a_s$ 的概率为

$$\Pr(y_i = a_s) = \frac{\exp(\nu_i^s)}{\sum_{t=1}^{S}\exp(\nu_i^t)} \tag{2.18}$$

这就是 2.3.3 节所介绍的多项对数概率模型。

**多项正态概率模型**(如 Daganzo,1979)代替的假定是,包含 $\epsilon_i^s$ 的向量具有一个方差 $\omega_s^2$ 和协方差 $\omega_{ss'}$ 的累积正态分布。考虑只有 3 个选项的简单情况,$S=3$,并定义 $v_i^{1\kappa} \equiv v_i^1 - v_i^\kappa$,且 $\epsilon_i^{1\kappa} \equiv \epsilon_i^1 - \epsilon_i^\kappa$。选择第一个选项 $a_1$ 的概率则变为

$$\Pr(y_i = a_1) = \int_{-\infty}^{\nu_i^{12}} \int_{-\infty}^{\nu_i^{13}} \varphi(\epsilon_i^{12}, \epsilon_i^{13}) \mathrm{d}\epsilon_i^{12} \mathrm{d}\epsilon_i^{13}$$

其中,$\varphi(\epsilon_i^{12}, \epsilon_i^{13})$ 是服从期望向量为 0,且协方差矩阵为

$$\Omega_1 = \begin{bmatrix} \omega_1^2 + \omega_2^2 - 2\omega_{12} & \\ \omega_1^2 - \omega_{13} - \omega_{12} + \omega_{23} & \omega_1^2 + \omega_3^2 - 2\omega_{13} \end{bmatrix}$$

的二元正态(bivariate normal)分布。注意这个选择的概率不能以闭式表达,而且需要 $S-1$ 个维度的积分。另一方面,多项正态概率模型允许效用是非独立的,而多项对数概率模型则相反。这因此释放了后一模型所具有的所谓的无关选项独立性(independence from irrelevant alternatives)(IIA)特征(参见 13.2 节)。

对于成对比较,如果单元 $i$ 更愿意选择选项 $s$ 而不是选项 $t$,则令 $y_{ist}=1$,否则为 0。在冈贝尔设定下,相应的概率变为

$$\Pr(y_{ist} = 1) = \frac{\exp(\nu_i^s - \nu_i^t)}{1 + \exp(\nu_i^s - \nu_i^t)}$$

其中,$v_i^1$ 通常设置为 0 以确保可辨识性。这一模型在布拉德和泰利(Bradley and Terry, 1952)以及鲁斯(Luce, 1959)之后,通常被称为布拉德-泰利-鲁斯模型(Bradley-Terry-Luce, BTL)。对于单元 $i$ 的一组成对比较的联合概率,通常被假定为这些概率的乘积。这一不太现实的独立性假设,可以通过引入潜变量而得到释放。

就排序而言,给定 $S$ 个选项当中的序列 $l$,令 $r_i^l$ 为选择项,并且设 $R_i \equiv (r_i^1, r_i^2, \cdots, r_i^s)$ 为单元 $i$ 的排序。一个排序的概率可以根据一个**效用序列**(utility ordering)进行解释,也可以根据 $S-1$ 个二元效用比较来进行表示,其中,第一个排序选项的效用大于第二个排序选项的效用,而第二选项的效用又要大于第三个选项的效用,依此类推。

对于服从冈贝尔分布的随机效用,就得到排序的对数概率模型(如 Luce, 1959; Plackett, 1975),即

$$\Pr(\mathbf{R}_i) = \prod_{\ell=1}^{S-1} \frac{\exp(\nu_i^{r_i^\ell})}{\sum_{m=\ell}^{s} \exp(\nu_i^{r_i^m})} \tag{2.19}$$

这一模型通常被称为激增**对数概率**(exploded logit)模型(Chapman and Staelin, 1982),因为排序概率是被写成一个依次的各个剩余选项的首选项概率的乘积。这一激增式的结果由鲁斯和萨普斯(luce and Suppes, 1965)以及伯格斯等人(Beggs et al., 1981)所证实。潜响应的视角揭示了激增对数概率能够直接被推导出来,而不需要构造这一行为假设,即选择过程是按顺序进行的。重要地,类似的分解在服从正态分布的效用中是不可获得的。冈贝尔模型是不可逆的,也就是说,从最差的选择开始的依次选择,可能会导致一个不同的排序概率。该模型的另一基本特征是无关选项独立性。

不同单元的不同选项集(alternative sets),如不同选区中符合条件的不同政党,可以简单地通过在上面的公式中用 $S_i$ 来代替 $S$ 而进行处理。当单元 $i$ 仅仅在全部选项集的一个子集中排序时,**部分排序**(Partial rankings)就出现了,如当实验设计被用来呈现不同单元特定的选项子集时(Durbin, 1951; Böckenholt, 1992)。通过设定选项集随着单元变动,这样的设计很容易进行操作。另一类型的部分排序是**顶级排序**(top-rankings),其中,

并非所有选项都被排序，而仅仅是 $P_i < S_i$ 的大部分首选项的子集。一个顶级排序的概率简单地就是等式(2.19)中第一项 $P_i$ 的乘积。注意，首选项的概率可以作为排序概率的一种特殊情况，即对于所有 $i$ 来说，当 $P_i = 1$ 时。

如果两个或者更多选项被给定同一个排序，它们就被称为是绑定(tied)。尽管绑定排序的概率在理论上是零，因为其效用是连续的，但是在实践中，绑定经常是可以观察到的。正如我们将要在2.5节看到，等式(2.19)与生存分析的考克斯回归中一个阶层(stratum)的偏似然具有相同的形式。开发这种二元性，我们可以利用以前的生存文献中已经提出的处理绑定的方法(参见2.5节)。

当研究选择行为时，收集比较响应数据是一种很自然的方法。在研究对象的普及性时，比较响应还可以有效地代替等级(rating)或温度计刻度的方法。刻度的使用可能需要借助于这一与事实不符的假设，即个体都是以相同的方式使用刻度(如 Brady, 1989, 1990)。然而，某些对象倾向于使用刻度的较高的边缘，而其他对象却使用较低的边缘。埃文和克罗斯尼克(Alwin and Krosnick, 1985)讨论了关于测量值的排序和等级设计的优缺点。

## 2.5 存续或生存的建模

对于某些事件来说，重要的响应可能是时间。在医学中，典型的例子是从一种疾病的发作或治疗的开始延续到死亡的**生存**时间(survival time)。在产品或零部件可靠性的研究中，比如日光灯，**失效**(failure)之前的使用寿命经常是调查出来的。代替使用这样的应用性特定术语，经济学家经常提到的是事件之间的**存续**。一般我们将延用这一术语，但偶尔也会偏离一点而使用生存时间或失效时间。

存续数据具有一些重要的区别性特征。存续总是正数，且某些存续并不明显地为人所知，因为该事件在观测期结束之前尚未发生(右删失)。此外，随着时间的消逝，协变量的值可能改变，协变量的效应也可能随着时间而改变。这些特征意味着，研究人员不能简单地对连续响应应用标准模型。

生存分析中另一种普遍的现象是左截断，或者叫延迟进入(delayed entry)，其中，当观测开始时，那些已经经历重要事件(如死亡)的单元被排除出研究(参见 Keiding, 1992)。注意，这一现象在计量经济学中被称为"跟踪库存抽样(stock sampling with follow up)"，因为仅仅那些在一个给定时间内处于"活着状态(alive state)"的单元会被抽中(如 Lancaster, 1990)。左删失与右截断要少见一些。左删失发生的条件是，仅仅知道在一个给定时间点——如观察开始——之前已经发生。右截断可能发生在回顾性的研究当中，如研究 AIDS 在已经发现疾病的病人身上的潜伏期。

在本节中，我们假定删失和截断过程是可以忽略的，因为在给定协变量的条件下，删失或截断的概率独立于事件发生的风险。我们也将把讨论限制在所谓的**吸收事件**(absorbing event)或状态，其中单元只能经历事件一次。复合事件的处理，比如复发性头疼，将推迟到第12章介绍，因为对同一单元所发生的复合事件之间的相关性必须进行考虑。简便起见，我们仅仅考虑单一的吸收事件，并且不对存在几种吸收状态的竞争风险模型(competing risk model)进行明确的论述。

存续模型通常不被定义为广义线性模型。然而，它证明广义线性模型经常能够适合收益似然(yield likelihoods)，这种似然与存续模型的收益似然是成比例的。

存续既可被认为是连续的时间，又可被认为是离散的时间，这些情形将在随后两节进行讨论。

### 2.5.1 连续时间存续

设 $T_i$ 为一个随机变量,代表一个单元 $i$ 的持续时间,从开始处于风险直到它经历这一事件或被删失。实现的存续用 $t_i$ 表示,有一个相应的指示变量(indicator variable)$\delta_i$,如果该单元经历了这个事件,取值为 1;如果该存续被删失了,则取值为 0。

单元 $i$ 的密度函数记为$f_i(t)$,累积分布函数成为

$$F_i(t) \equiv \int_0^t f_i(u)\,du$$

这一生存函数,存续超过 $t$ 的概率就可以定义为

$$S_i(t) \equiv 1 - F_i(t)$$

**风险**(hazard),有时也称为**发生率**(incidence rate)或**瞬时风险**(instantaneous risk),可以定义为

$$h_i(t) \equiv \lim_{\Delta \to 0}\left\{\frac{\Pr(t \leqslant T_i < t + \Delta \mid T_i \geqslant t)}{\Delta}\right\}$$

不严格地说,这是单元 $i$ 的一个事件在时间 $t$ 上的"风险",给定条件是该事件尚未发生且该单元因此仍然"处于风险中"。

它是根据这些定义得出的,即

$$h_i(t) = \frac{f_i(t)}{S_i(t)} = -\frac{\partial \ln S_i(t)}{\partial t} \tag{2.20}$$

既然 $S_i(0) = 1$,那么它也服从

$$S_i(t) = \exp\left[-\int_0^t h_i(u)\,du\right]$$

定义**综合风险**(integrated hazard)或**累积风险**(cumulative hazard)为

$$H_i(t) \equiv \int_0^t h_i(u)\,du$$

单元 $i$ 对于似然的贡献为

$$l_i = S_i(t_i)h_i(t_i)^{\delta_i} = \exp[-H_i(t_i)]h_i(t_i)^{\delta_i} \tag{2.21}$$

**加速失效时间模型(Accelerated failure time model)**

设定存续模型的一种方法是

$$\ln T_i = \beta_0 + \nu_i + \epsilon_i$$

其中,$\nu_i$ 并不包含一个常数,或可以表示为另一种形式:

$$T_i = \exp(\beta_0)\exp(\nu_i)\exp(\epsilon_i)$$

我们看到,协变量产生的效果在时间刻度上是直接倍增的,从而加速或减缓着存续。此类加速失效时间模型的生存函数因此可以表示为

$$S_i(t) = S_0(\exp(\nu_i)t)$$

其中,$S_0(t)$是 $\boldsymbol{x}_i = \boldsymbol{0}$ 的基线生存函数。加速失效时间模型的范例包括**对数正态存续模型**(log-normal duration model),其中,$\epsilon_i$ 服从正态分布;对数-对数概率**存续模型**(log-logistic duration model),带有一个丢数概率分布的 $\epsilon_i$;以及**威布尔存续模型**(Weibull duration model),如果 $\epsilon_i$ 服从冈贝尔分布,参见前一节中的式(2.17)。

**比例风险模型(Proportional hazards model)**

该风险可以模型化为

$$h_i(t) = h^0(t)\exp(\nu_i) \tag{2.22}$$

其中,$h^0(t)$是"基线"风险,即当所有协变量为 0 时的风险(该线性估计量不包含一个常数)。考虑两个单元 $i$ 和 $i'$,我们得到

$$\frac{h_i(t)}{h_{i'}(t)} = \exp(\nu_i - \nu_{i'}) \tag{2.23}$$

任何两个单元的风险函数都与时间成比例。

上面介绍的威布尔存续模型是独特的，因为它同时拥有比例风险和加速失效时间的特性（如 Cox and Oakes，1984）。**指数存续模型**（exponential duration model）是威布尔模型的特例，其中，基线风险是固定的，$h^0(t) = h^0$。这一特征在逐段指数存续模型中被释放，其中，基线风险函数被假定为在时间间距 $s$ 上是逐段固定的，即对于 $\tau_{s-1} \leqslant t < \tau_s, s = 1, 2, \cdots, S, h^0(t) = h_s$。

有趣的是，已经证实当数据适当扩展时，逐段指数存续模型的似然是与泊松模型的似然是成比例的（Holford，1980）。设 $\theta_i = \exp(\nu_i)$，则单元 $i$ 有风险 $h_s\theta_i$。对于在某个区间中被删失或者失败的单元，它对于似然的贡献变为（参见式（2.21））

$$l_i = \underbrace{(h_s\theta_i)}_{h_i(t_i)}{}^{\delta_i}\exp(-\underbrace{\sum_{r=1}^{s} h_r\theta_i d_{ir}}_{H_i(t_i)}) \tag{2.24}$$

其中，$d_{ir}$是单元 $i$ 在区间 $r$ 中所花的时间，$d_{ir} = \min(t_i, \tau_r) - \tau_{r-1}$。这样就可以写为

$$l_i = \prod_{r=1}^{s} (h_r\theta_i)^{y_{ir}}\exp(-h_r\theta_i d_{ir}) \tag{2.25}$$

其中，当 $r < s$ 且 $y_{is} = \delta_i$ 时，$y_{ir} = 0$。这与均值为 $h_r\theta_i d_{ir}$的 $S$ 个独立泊松变量 $y_{ir}$的似然贡献是成比例的，参见式（2.3），也可参阅 Clayton（1988）。

我们用若干观测数据（或“风险集”）——等于在该单元失效（或删失）时间之前的或包括失效时间的区间数——来表示每一个单元，正如展示 2.1 上半部分所示。然后，这个模型就可以通过采用对数链接的泊松回归进行拟合，将 $y_{ir}$视为响应变量，$\ln(d_{ir})$作为一个偏移量（一个回归系数设置为 1 的协变量），区间的虚拟变量作为解释变量，来考虑不同的逐段固定风险 $h_r$。显然

$$y_{ir} \sim \text{Poisson}(\mu_{ir})$$

其中

$$\ln(\mu_{ir}) = \nu_{ir}$$

且

$$\nu_{ir} = \ln(d_{ir}) + \ln(h_r) + \mathbf{x}_s'\boldsymbol{\beta}$$

因此，建构生存模型的一种途径是，将跟踪期（follow-up period）划分成区间，通过这种方法，风险就可以被假定为逐段固定的，并且可以运用泊松回归。布里斯罗和戴伊（Breslow and Day，1987：137）指出了泊松回归如何应用于同一性链接和幂数链接（identity and power links）。假定一个逐段线性对数风险（piecewise linear log hazard）符合一个逐段龚柏兹（Gompertz）分布（如 Lillard，1993）。

另一种方法是对众多区间进行界定，每一个区间都存在多个独特的失效时间，开始于（紧随其后）一个独特失效时间，并且结束于（紧随其后）下一个大的独特失效时间（Holford，1976；Whitehead，1980；Laird and Olivier，1981）。这就相当于著名的**考克斯比例风险模型**，因为“饱和模型”或非参数模型——每个风险集都有单独的常数——被应用于基准风险的估计。

用包含对数链接的泊松回归来估计考克斯模型所必需的数据扩展，如展示 2.1 下半部分中的人工处理的数据集所示。可以推出，相应的似然是与考克斯回归的偏似然成比例的，即

$$lp = \prod_r \frac{\exp(\nu_{(r)})}{\sum_{i\in R(t_{(r)})}\exp(\nu_i)}$$

其中,$\nu_{(r)}$是单元在第 $r$ 个有序失效时间 $t_{(r)}$ 时失效的线性估计量,$R(t_{(r)})$是第 $r$ 个失效时间的风险集(参见展示 2.1,其中失效和贡献分子的单元被包括在一个小方框内)。这种偏似然可以通过采用一种剖面似然方法(profile likelihood approach)来估计基线风险而得到(Johansen,1983)。必须注意,这个偏似然等同于带有特定成对截距(pair-specific intercepts)的 logistic 回归的条件似然,广泛地应用于病例对照研究(matched case-control studies)。

**展示2.1 连续的存续:原始数据和扩展数据**

**逐段指数模型**

原始数据

| $i$ | $t_i$ | $\delta_i$ |
|---|---|---|
| 1 | 5.2 | 0 |
| 2 | 4.5 | 1 |

$\tau_0=0$, $\tau_1=1.5$, $\tau_2=5$, $\tau_3=8$ →

泊松回归

| $i$ | $r$ | $\tau_r$ | $d_{ir}$ | $y_{ir}$ |
|---|---|---|---|---|
| 1 | 1 | 1.5 | 1.5 | 0 |
| 1 | 2 | 5 | 3.5 | 0 |
| 1 | 3 | 8 | 0.2 | 0 |
| 2 | 1 | 1.5 | 1.5 | 0 |
| 2 | 2 | 5 | 3 | 1 |

**考克斯比例风险模型**

原始数据

| $i$ | $t_i$ | $\delta_i$ |
|---|---|---|
| 1 | 5.6 | 0 |
| 2 | 4.2 | 1 |
| 3 | 2.8 | 1 |
| 4 | 5.1 | 1 |

→ 泊松回归

| $i$ | $r$ | $t_{(r)}$ | $d_{ir}$ | $y_{ir}$ |
|---|---|---|---|---|
| 1 | 1 | 2.8 | 2.8 | 0 |
| 1 | 2 | 4.2 | 1.4 | 0 |
| 1 | 3 | 5.1 | 0.9 | 0 |
| 2 | 1 | 2.8 | 2.8 | 0 |
| 2 | 2 | 4.2 | 1.4 | 1 |
| 3 | 1 | 2.8 | 2.8 | 1 |
| 4 | 1 | 2.8 | 2.8 | 0 |
| 4 | 2 | 4.2 | 1.4 | 0 |
| 4 | 3 | 5.1 | 0.9 | 1 |

→ 风险集

| $r$ | $t_{(r)}$ | $i\in R(t_{(r)})$ |
|---|---|---|
| 1 | 2.8 | 1 |
| 2 | 2.8 | 2 |
| 3 | 2.8 | [3] |
| 1 | 2.8 | 4 |
| 2 | 4.2 | 1 |
| 2 | 4.2 | [2] |
| 2 | 4.2 | 4 |
| 3 | 5.1 | 1 |
| 3 | 5.1 | [4] |

在考克斯回归中,对每一个失效都允许有一个不同的基线风险值,与此不同的是,运用泊松回归方法,基线风险可以被模型化为一个平滑时间函数(smooth function of time)。一个时间的多项式可以应用于对数风险、选择性的样条(splines)或者分数多项式。仅当 $\ln(d_{ir})$作为一个残差被包括进去,这个估计函数才可以被理解为基线风险函数,这一点为戈德斯坦恩(Goldstein,2003,第 10 章)所忽视。必须注意,尽管风险被模型化为一个关于失效时间 $t_{(r)}$ 的平滑函数,但是它在相邻的失效时间之间是固定的。

实践中经常会出现绑定的或同一的存续(tied or identical durations),尽管与连续时间存续模型不相一致。处理绑定的适当方法,是将对绑定存续所有可能的排列的似然贡献

进行求和(如 Kalbfleisch and Prentice,2002),但是这样会变得非常复杂。因此,可以采用近似的方法,其中最广泛使用的是 Peto-Breslow 方法(Peto,1972;Breslow,1974)。该方法是假设无论哪一个单元失效,所有绑定单元仍然处于风险之中。只要绑定的数量是适中的,Peto-Breslow 方法看来就是十分有效的(Farewell and Prentice,1980)。艾弗朗(Efron,1977)提出了一个更好的近似方法,其中,绑定单元在分母中的贡献接连地降低权重(downweight),以反映不断减少的风险集。当存在很多绑定时,更适合将存续视为离散的而非连续的。

## 2.5.2 离散时间存续

离散时间存续的数据大多数来自于持续时间过程中的区间删失。另一来源是离散时间过程,其中,事件只能发生于离散的时间点上,如就选举次数而言的政党忠诚度。在两者中的任一情况下,将离散存续模型化为区间删失都是非常有用的。设 $\tau_s$ 为删失界限,那么我们已知

$$\tau_{s-1} \leqslant T_i < \tau_s$$

另外,离散生存时间可能是右删失的,即

$$\tau_{s-1} < T_i$$

**比例比率模型**

2.3.4 节介绍的针对定序响应的比例比率模型,也可以用来建立离散时间存续的模型。$T_i$ 小于 $\tau_s$ 的概率为

$$\Pr(T_i < \tau_s) = \frac{\exp(\boldsymbol{x}_i'\boldsymbol{\beta} + \kappa_s)}{1 + \exp(\boldsymbol{x}_i'\boldsymbol{\beta} + \kappa_s)}$$

而生存时间位于第 $k$ 个区间 $\tau_{s-1} \leqslant T_i < \tau_s$ 中的概率为

$$\Pr(\tau_{s-1} \leqslant T_i < \tau_s) = \Pr(T_i < \tau_s) - \Pr(T_i < \tau_{s-1})$$

其中,$\Pr(T_i < \tau_0)0$。需要注意的是,这里的 $\beta$ 与 2.3.4 节中的 $\boldsymbol{\beta}$ 具有相反的意义,因而大的系数意味着增加的风险。在没有出现右删失的情况下,这是生存时间位于第 $s$ 个区间的所有观测数据的似然贡献。对于在第 $s$ 个区间后删失的观测数据,似然贡献是 $\Pr(T_i \geqslant \tau_s) = 1 - \Pr(T_i < \tau_s)$。比例比率模型被本尼特(Bennett,1983)、汉和豪斯曼(Han and Hausman,1990)、埃兹泽特和怀特海德(Ezzet and Whitehead,1991)以及海德克等人(Hedeker et al. ,2000)应用于离散生存时间数据。

比例比率模型也可以被给予一个潜响应解释(参见 2.4 节)

$$y_i^* = \boldsymbol{x}_i'\boldsymbol{\beta} + \epsilon_i \tag{2.26}$$

其中,$\epsilon_i$ 具有一个 logistic 分布。如果假定 $\epsilon_i$ 服从标准正态分布,那么就得到定序正态概率模型。当 $\kappa_{s-1} \leqslant y_i^* < \kappa_s$,事件就发生在第 $s$ 个区间中,即

$$\Pr(\tau_{s-1} \leqslant T_i < \tau_s) = \Pr(\kappa_{s-1} \leqslant y_i^* < \kappa_s)$$

潜响应 $y_i^*$ 可以因此被认为是 $T_i$ 的单调转换(monotonic transformation),因而 $y_i^* = \kappa_s$ 相当于 $T_i = \tau_s$。通过将阙值参数 $\kappa_s$ 约束为等间距,从而使 $T_i$ 向 $y_i^*$ 的转化是线性的,并可以估计关于 $T_i$ 的线性回归模型的适当性。

**基于离散时间风险的模型**

第 $s$ 个区间的离散时间风险被定义为事件在第 $s$ 个区间中发生的概率,假定它还没有发生,则

$$h_i(s) \equiv \Pr(\tau_{s-1} \leqslant T_i < \tau_s \mid T_i \geqslant \tau_{s-1}) = \frac{\Pr(\tau_{s-1} \leqslant T_i < \tau_s)}{\Pr(T_i \geqslant \tau_{s-1})}$$

生存时间位于第 $s$ 个区间的一个单元，它的似然贡献为

$$l_i = h_i(s)\prod_{r=1}^{s-1}[1-h_i(r)] = \prod_{r=1}^{s}h_i(r)^{y_{ir}}[1-h_i(r)]^{(1-y_{ir})} \quad 且 \quad y_{is}=1 \qquad (2.27)$$

其中，$y_{ir}$是一个指示变量，当事件发生在第 $r$ 个区间时，它等于 1；否则为 0。也就是说，当 $r<s$时，$y_{ir}=0$；当 $r=s$ 时，$y_{ir}=1$。在第 $k$ 个区间后删失的单元，其似然贡献具有同一形式，即

$$l_i = \prod_{r=1}^{s}[1-h_i(r)] = \prod_{r=1}^{s}h_i(r)^{y_{ir}}[1-h_i(r)]^{1-y_{ir}} \quad 且 \quad y_{is}=0 \qquad (2.28)$$

删失和非删失观测数据的似然贡献，都正好是 $s$ 个独立二项响应 $y_{ir}$——具有伯努利概率 $h_i(r)$，$r=1,\cdots,s$——的似然贡献。因此，对于一个要么失败，要么在第 $s$ 个区间删失的单元，我们将数据扩展到 $s$ 个记录，并且构建指示变量 $y_{ir}$，如展示 2.2 所示。这样，离散时间生存模型就可以被写成二项响应的广义线性模型。

一个可能性是运用 logistic 回归，其中对每一个区间都有一个单独的常数，即

$$\ln\frac{h_i(r)}{1-h_i(r)} = \nu_i + \kappa_r \qquad (2.29)$$

这一由考克斯(Cox，1972)提出的模型，经常被视认为是一种比例比率模型。然而，尽管在这里比例被应用于事件发生在一个区间中的条件比率(假定它还没有发生)，但比例比率模型中的比例，如前面的章节所介绍的，是应用于事件发生在一个给定区间或更早的比率。除了线性估计量的符号，上述离散时间生存数据的对数模型等价于 2.3.4 节介绍的升级比模型。升级比模型对于连续的过程是非常有用的，其中阶段(如教育达到的水平)不能被跳过，且关注点集中在(不)连续的超过一个阶段的比率(假定这个阶段已经达到)。关于这个模型的介绍参见 Jenkins(1995)、Singer and Willett(1993)。

**展示2.2 离散时间存续：原始数据和扩展数据**

原始数据

| $i$ | $s$ | $\delta_i$ |
|---|---|---|
| 1 | 3 | 0 |
| 2 | 2 | 1 |
| 3 | 1 | 1 |
| 4 | 3 | 1 |

→

对数概率回归或余重对数回归

| $i$ | $r$ | $y_{ir}$ |
|---|---|---|
| 1 | 1 | 0 |
| 1 | 2 | 0 |
| 1 | 3 | 0 |
| 2 | 1 | 0 |
| 2 | 2 | 1 |
| 3 | 1 | 1 |
| 4 | 1 | 0 |
| 4 | 2 | 0 |
| 4 | 3 | 1 |

如果一个考克斯比例风险模型被应用于未观测到的连续生存时间和观测到的离散生存时间(它们都被视为是区间删失的)，当这种离散时间风险应用一个余重对数链接时，可以证明其似然贡献等同于方程式(2.27)和方程式(2.28)中的似然贡献(如 Thompson，1977)，即

$$\ln\left\{-\ln[1-h_i(r)]\right\}=\nu_i+\kappa_r \tag{2.30}$$

## 2.6 总结与进一步阅读

我们已经介绍了广泛的响应过程,包括连续、二项、分组、删失、定序和比较响应,同时还有事件数以及离散和连续时间的存续。大多数响应过程或多或少地能够直接表示为广义线性模型,且很多可以表示为潜响应模型。这些响应过程的模型,为后续章节所介绍的更多广义模型奠定了基础。而且,本书第9—14章的应用都是根据响应过程的类型来进行组织的。

关于广义线性模型,一些有用的著作包括 McCullagh and Nelder(1989)、Aitkin et al.(1989,2004)、Fahrmeir and Tutz(2001)。

对于分类响应,我们推荐 Long(1997)、Agresti(1996,2002)、Collett(2002)、Andersen(1980)。关于定序响应的讨论请见 Clogg and Shihadeh(1994)、Johnson and Albert(1999),关于事件数的讨论请参考 Cameron and Trivedi(1998)、Winkelmann(2003)。马达拉(Maddala,1983)讨论了分类响应和有限依赖响应,如删失和截断响应。

下列著作是围绕比较响应的,关注排序的是 Marden(1995),关注成对比较的是 David(1988),关注多项类响应或离散选项的是 Train(1986,2003)。

关于存续或"生存"的模型建构,一些有用的书籍包括如下(按照难度升序排列):Singer and Willett(2003)、Allison(1984)、Hosmer and Lemeshow(1999)、Collett(2003)、Breslow and Day(1987)、Klein and Moeschberger(2003)、Therneau and Grambsch(2000)、Vermunt(1997)、Cox and Oakes(1984)、Kalbfleisch and Prentice(2002)、Andersen et al.(1993)。

# 3 经典潜变量模型

## 3.1 介　绍

本章我们研究经典的潜变量模型。描述这些模型的作用有 3 点：首先，给出一个关于潜变量传统上是如何应用于各类统计学分支的概要；其次，使读者熟悉将要出现在本书后面部分中的基本观点、符号和术语；最后，着手统一不同的方法，为第 4 章讨论一般模型框架做出准备。

本章考虑的潜变量模型包括：

- 多层次回归模型。
- 因子模型。
- 题器响应模型。
- 结构方程模型。
- 潜类模型。
- 纵贯性数据模型。

很明显，潜变量模型由于各种目的而被广泛应用于不同的理论学科，这将在本书的应用性章节中多次涉及。而且，其中一部分模型被应用于交叉学科领域，尽管其他模型在专业学科之外几乎无人知晓。

## 3.2 多层次回归模型

当很多单元嵌套在群组当中时，多层次数据就出现了。例如，教室中的学生、医院里的病人、个人的左右眼睛等。我们将基本的单元（如学生、病人或眼睛）归为第一层次的单元，将群组（如班级、医院或头）归为第二层次的单元。如果这些群组本身再次聚集到"更高层次"（超级的）群组，例如，学生嵌套在班级中，班级再次嵌套在学校中，那么数据就具有三层次结构。多层次建模的重要发展起始于学校体系的研究；请参见如 Aitkin et al.（1981）、Aitkin and Longford（1986）。

属于同一群组的单元共享着相同的特定群组影响。例如，由同一个老师授课的同一班级学生，以及被他们的父母送到同一所学校的学生（通过选择或由于居住位置）。然而，在一项分析当中，我们不能指望像协变量一样包含所有的特定群组影响。这是因为我们通常对相应的协变量了解有限，而且我们的数据集也缺少这些协变量的信息。结果

就存在群组层次的未观测到的异质性，导致在条件依赖于协变量之后，同一群组中单元的响应之间产生相关性。这在 1.4 节的图 1.4 中已进行了说明。

在多层次回归中，未观测到的异质性是通过在多元回归模型中纳入随机效应来建立模型。随机效应有两种类型：随机截距与随机系数。随机截距反映的是全部响应中的未观测到的异质性，而随机系数反映的是解释变量对响应变量的效应中的未观测到的异质性。

### 3.2.1 双层次随机截距模型

将第二层次的单元（如学校）标记为 $j=1,\cdots,J$，第一层次的单元（如学生）标记为 $i=1,\cdots,n_j$。考虑单含一个特定学生协变量 $x_{ij}$ 的双层次随机截距模型

$$y_{ij} = \eta_{0j} + \beta_1 x_{ij} + \epsilon_{ij} \tag{3.1}$$

其中，$\eta_{0j}$ 是特定学校的截距，$\beta_1$ 是回归系数，$\epsilon_{ij}$ 是第一层次的残差项。$\eta_{0j}$ 可被模型化为

$$\eta_{0j} = \gamma_{00} + \zeta_{0j} \tag{3.2}$$

其中，$\gamma_{00}$ 是平均截距，而 $\zeta_{0j}$ 是特定学校的截距 $\eta_{0j}$ 与均值的偏差。定义 $\theta \equiv \text{Var}(\epsilon_{ij})$，$\psi \equiv \text{Var}(\zeta_{0j})$，通常假定群组 $j$ 是独立的，并且

$$\epsilon_{ij} \mid x_{ij} \sim \text{N}(0,\theta)$$
$$\text{Cov}(\epsilon_{ij},\epsilon_{i'j}) = 0 \quad i \neq i'$$
$$\zeta_{0j} \mid x_{ij} \sim \text{N}(0,\psi)$$
$$\text{Cov}(\zeta_{0j},\epsilon_{ij}) = 0$$

注意第 1 个假设意味着 $\epsilon_{ij}$ 与协变量 $x_{ij}$ 是不相关的，第 3 个假设则意味着 $\zeta_{0j}$ 与 $x_{ij}$ 是不相关的。当这些假设被规定后，稍不注意，对 $x_{ij}$ 的条件依赖通常会被忽略。从现在开始，我们将坚持这一约定，牢记随机项的期望值应当被解释为对协变量的条件依赖。

将 $\eta_{0j}$ 的第二层次的模型式(3.2)代入到 $y_{ij}$ 的第一层次的模型式(3.1)中，可得到该模型的**简化形式**，即

$$y_{ij} = \gamma_{00} + \zeta_{0j} + \beta_1 x_{ij} + \epsilon_{ij} \tag{3.3}$$

这个模型类似于传统的协方差分析（ANCOVA）模型，但是，$\zeta_{0j}$ 是学校“因子”的随机效应，而不是学校的固定效应。在方差分析（ANOVA）的术语中，学校被认为是一个**随机因子**（不能与 3.3.2 节讨论的“公共因子”相混淆）。通过为截距假设一种分布，学校效应可经由一个单一的参数（即方差 $\psi$）而获得，而不是每个学校（除了只有一个学校）都有一个单独的参数。如果我们希望在数据集中对学校的总体而不是某个特定的学校作出推论，那么将学校视为一个随机因子就是非常合适的。我们可参阅本章 3.6.1 节和“估计”一章（第 6 章）中 6.1 节和 6.10 节所涉及的固定效应和随机效应之间的区别。随机因子学校与固定因子之间的交互作用产生随机系数模型，这将在 3.2.2 节中进行讨论。

随机截距模型是**混合效应模型**或**线性混合模型**的一个实例，因为它既包括固定效应 $\gamma_{00}$ 和 $\beta_1$，又包括随机效应 $\zeta_{0j}$ 和残差 $\epsilon_{ij}$。我们可将简化的方程式划分为固定部分和随机部分，即

$$y_{ij} = \underbrace{\gamma_{00} + \beta_1 x_{ij}}_{\text{固定部分}} + \underbrace{\zeta_{0j} + \epsilon_{ij}}_{\text{随机部分}}$$

其中，随机部分的各项之和可被认为是一个总残差，即 $\xi_{ij} = \zeta_{0j} + \epsilon_{ij}$。由于存在这个误差成分，该模型有时也称为误差成分模型（error-component model）。这个总残差的方差，或者与其等价的给定 $x_{ij}$ 的响应的条件方差是

$$\text{Var}(y_{ij} \mid x_{ij}) = \text{Var}(\zeta_{0j} + \epsilon_{ij}) = \psi + \theta$$

该方差由两个**方差成分**组成,学校间方差 $\psi$ 和学校内方差 $\theta$。如果忽略式(3.1)中的 $x_{ij}$,则该模型可被称为一个**方差成分模型**(variance components model)。

给定随机截距 $\zeta_{0j}$ 和协变量值 $x_{ij}, x_{i'j}$,则同一个第二层次单元中的任意两个响应 $y_{ij}$ 和 $y_{i'j}$ 是条件独立的,即

$$\mathrm{Cov}(y_{ij}, y_{i'j} \mid \zeta_{0j}, x_{ij}, x_{i'j}) = \mathrm{Cov}(\epsilon_{ij}, \epsilon_{i'j}) = 0, \quad i' \neq i$$

但是,由于在同一所学校中,随机截距是由学生们所共享的,因此,它们导致在条件依赖于协变量后,同一所学校内学生的响应之间存在相关性。这种相关性通常是以一个群组内相关的形式表示出来,即所谓的**组内相关**(intraclass correlation)。这个组内相关 $\rho$ 写为

$$\rho \equiv \mathrm{Cor}(y_{ij}, y_{i'j} \mid x_{ij}, x_{i'j}) = \mathrm{Cor}(\zeta_{0j} + \epsilon_{ij}, \zeta_{0j} + \epsilon_{i'j}) = \frac{\psi}{\psi + \theta}, \quad i' \neq i$$

因而组内相关表现为总残方差 $\psi + \theta$ 的比例,即是由学校间残方差 $\psi$ 造成的。

在学校层次的模型(3.2)中,我们可尝试着通过纳入学校层次的协变量 $\omega_j$(如教师学生比)来"解释"学校间的变异性,即

$$\eta_{0j} = \gamma_{00} + \gamma_{01} w_j + \zeta_{0j} \tag{3.4}$$

这里,$\gamma_{00}$ 和 $\gamma_{01}$ 是固定系数,而 $\zeta_{0j}$ 现在变成一个学校层次的残差或干扰项。将这一模型代入式(3.1),我们得到简化形式

$$y_{ij} = \gamma_{00} + \gamma_{01} w_j + \zeta_{0j} + \beta_1 x_{ij} + \epsilon_{ij} \tag{3.5}$$

**内生性**(Endogeneity)

随机截距模型的一个假设是 $\mathrm{E}(\zeta_{0j} \mid x_{ij}) = 0$,这意味着 $\mathrm{Cov}(\zeta_{0j}, x_{ij}) = 0$。如果随机截距表示的是缺失协变量的效应,那么这个假设就会常常被违反,因为这些缺失协变量完全可能与观测到的协变量相关。在这种情况下,观测到的协变量则被称为**内生的**(endogenous);参见恩格等人(Engle et al.,1983)关于内生性和外生性的讨论。

我们将讨论这一情形,其中设定的模型为

$$y_{ij} = \gamma_{00} + \zeta'_{0j} + \beta_1 x_{ij} + \epsilon_{ij} \tag{3.6}$$

但正确的模型是式(3.5)所表示的。换句话说,我们略去了一个群组层次的协变量 $\omega_j$。如果随机截距是表示省略掉的协变量效应,那么它应该有期望值 $\gamma_{01}\omega_j$,可记为

$$\zeta'_{0j} = \gamma_{01}\omega_j + \zeta'_{0j} \tag{3.7}$$

如果 $\omega_j$ 和 $x_{ij}$ 是相关的,那么省略 $\omega_j$ 就有问题。这种相关性可被表示为下述回归:

$$\omega_j = \alpha_0 + \alpha_1 \bar{x}_{.j} + u_j$$

其中,$\bar{x}_{.j}$ 是 $x_{ij}$ 的群组均值。注意,我们是在群组均值 $\bar{x}_{.j}$ 而不是 $x_{ij}$ 上对 $\omega_j$ 进行回归,因为 $x_{ij} = (x_{ij} - \bar{x}_{.j}) + \bar{x}_{.j}$ 并且 $\omega_j$ 在 $(x_{ij} - \bar{x}_{.j})$ 上的回归系数是零。

将 $\omega_j$ 代入式(3.7),可看到随机截距依赖于 $\bar{x}_{.j}$,即

$$\begin{aligned}\zeta'_{0j} &= \underbrace{\gamma_{01}\alpha_0}_{\delta_0} + \underbrace{\gamma_{01}\alpha_1}_{\delta_1}\bar{x}_{.j} + \underbrace{\gamma_{01}u_j + \zeta_{0j}}_{\overline{\omega}_j} \\ &= \delta_0 + \delta_1 \bar{x}_{.j} + \overline{\omega}_j\end{aligned}$$

将 $\zeta'_{0j}$ 代入式(3.6),得到模型的简化形式

$$y_{ij} = \gamma_{00} + \delta_0 + \delta_1 \bar{x}_{.j} + \beta_1 x_{ij} + \overline{\omega}_j + \epsilon_{ij} \tag{3.8}$$

在正确设定的模型中,尽管我们省略了群组层次的协变量 $\omega_j$,但是通过纳入群组均值 $\bar{x}_{.j}$ 作为一个单独的协变量,$x_{ij}$ 的系数则变成必要的参数 $\beta_1$。然而,如果 $\delta_1 = \gamma_{01}\alpha_1 \neq 0$,那么在模型中省略 $\bar{x}_j$ 将会产生一个对 $\beta_1$ 的有偏估计。

有必要重新回顾1.4节简要讨论的假设例子。其中,单元 $i$ 是人,群组 $j$ 是国家。响

应 $y_{ij}$是生命的长度,$x_{ij}$是红肉消耗量的显露,而 $\omega_j$ 是国家生活水平的某种指标。国家层次的随机截距,它表示的是不同国家的生命期望之间的未知变异性,能完全包含省略变量(即生活水平 $\omega_j$)的效应,省略的这个变量是依次与平均的红肉消耗量 $\bar{x}_{.j}$相关的。我们可假设 $\gamma_{01}>0,\alpha_1>0$,于是 $\delta_1>0$。国家均值 $\bar{x}_{.j}$对特定国家截距的正效应可以在 1.4 节的图 1.7 中看到。使用设定错误的模型(不纳入 $\bar{x}_{.j}$作为一个单独的协变量),红肉对生命期望所产生的真实负效应会因此被低估,原因是省略掉的协变量生活水平,导致了红肉与生命期望之间在国家层次上的正关系。

式(3.8)中的模型可以用另一种方式写成

$$y_{ij} = \gamma_{00} + \delta_0 + (\delta_1 + \beta_1)\bar{x}_{.j} + \beta_1(x_{ij} + \bar{x}_{.j}) + \bar{\omega}_j + \epsilon_{ij}$$

其中,$\delta_1+\beta_1$ 是群组间效应,$\beta_1$ 是群组内效应(这一参数化设置可能是更可取的,因为协变量 $\bar{x}_{.j}$和 $x_{ij}-\bar{x}_{.j}$是不相关的)。关于群组间和群组内回归效应的等价性的Wald检验(也就是零假设 $\delta_1=0$ 的一种检验),等同于随机截距模型的豪斯曼设定检验(Hausman specification test)(如 Hausman,1978);也可参见 8.5.1 节。有些经济学家认为,一个显著的豪斯曼检验意味着必须抛弃随机截距模型而支持一种固定效应模型。然而,这是一种误导,因为如上文所述,只要除 $x_{ij}$之外,群组均值 $\bar{x}_{.j}$也被纳入作为协变量,$\beta_1$ 就可以被无偏估计。可以参考斯奈德和伯克霍夫(Snijders and Berkhof,2004)的深入讨论,其中也展示了下一节所描述的几种随机效应模型的类似结果。

注意,第一层次的残差 $\epsilon_{ij}$和协变量之间的相关性仍然遗留下一个潜在的问题,如同在任何线性回归模型中一样。遗憾的是,这一问题并不像随机截距那样容易被发现和克服。

### 3.2.2 双层次随机系数模型

通过允许一个协变量 $x_{ij}$的斜率在群组之间(如学校)变动,双层次随机截距模型可以扩展为一个**随机系数模型**。扩展模型可被表示为

$$y_{ij} = \eta_{0j} + \eta_{1j}x_{ij} + \epsilon_{ij} \tag{3.9}$$

其中,$x_{ij}$是一个学生层次的协变量,如性别,$\eta_{0j}$和 $\eta_{1j}$分别是第 $j$ 个学校的截距和斜率。截距的学校间变异性的建模是下面的第一个等式,同时我们也给出一个关于斜率的类似模型,即

$$\begin{aligned} \eta_{0j} &= \gamma_{00} + \gamma_{01}\omega_j + \zeta_{0j} \\ \eta_{1j} &= \gamma_{10} + \gamma_{11}\omega_j + \zeta_{1j} \end{aligned} \tag{3.10}$$

学校层次的残差或干扰项 $\zeta_{0j},\zeta_{1j}$都被规定为具有零均值、方差 $\psi_1$ 和 $\psi_2$ 以及协变量 $\psi_{21}$的二元正态分布。

将式(3.10)中系数 $\eta_{0j}$和 $\eta_{1j}$的学校层次模型代入式(3.9)中的学生层次模型,得到简化形式

$$\begin{aligned} y_{ij} &= \underbrace{\gamma_{00} + \gamma_{01}\omega_j + \zeta_{0j}}_{\eta_{0j}} + \underbrace{(\gamma_{10} + \gamma_{11}\omega_j + \zeta_{1j})}_{\eta_{1j}}x_{ij} + \epsilon_{ij} \\ &= \gamma_{00} + \gamma_{01}\omega_j + \gamma_{10}x_{ij} + \gamma_{11}(\omega_j x_{ij}) + \zeta_{0j} + \zeta_{1j}x_{ij} + \epsilon_{ij} \end{aligned}$$

与随机截距模型相反,随机系数模型导致异方差的响应,原因是条件方差

$$\mathrm{Var}(y_{ij} \mid x_{ij},\omega_j) = \psi_1 + 2\psi_{21}x_{ij} + \psi_2 x_{ij}^2 + \theta \tag{3.11}$$

依赖于 $x_{ij}$。在这种情况下,组内相关也依赖于协变量。

必须注意的是,随机截距方差以及截距和斜率之间的相关性,对 $x_{ij}$的转化来说并不是恒定的。如图 3.1 所示,同样的特定群组的回归线展示在两个面板图中,但是右图中

的解释变量 $x'_{ij}=x_{ij}-3.5$ 相应地转化为左图中的解释变量 $x_{ij}$。截距是回归线与0上垂直线的交叉。很明显,左图中截距的变动程度要大于右图,但是左图中截距和斜率的相关是负的,而在右图中则是正相关。

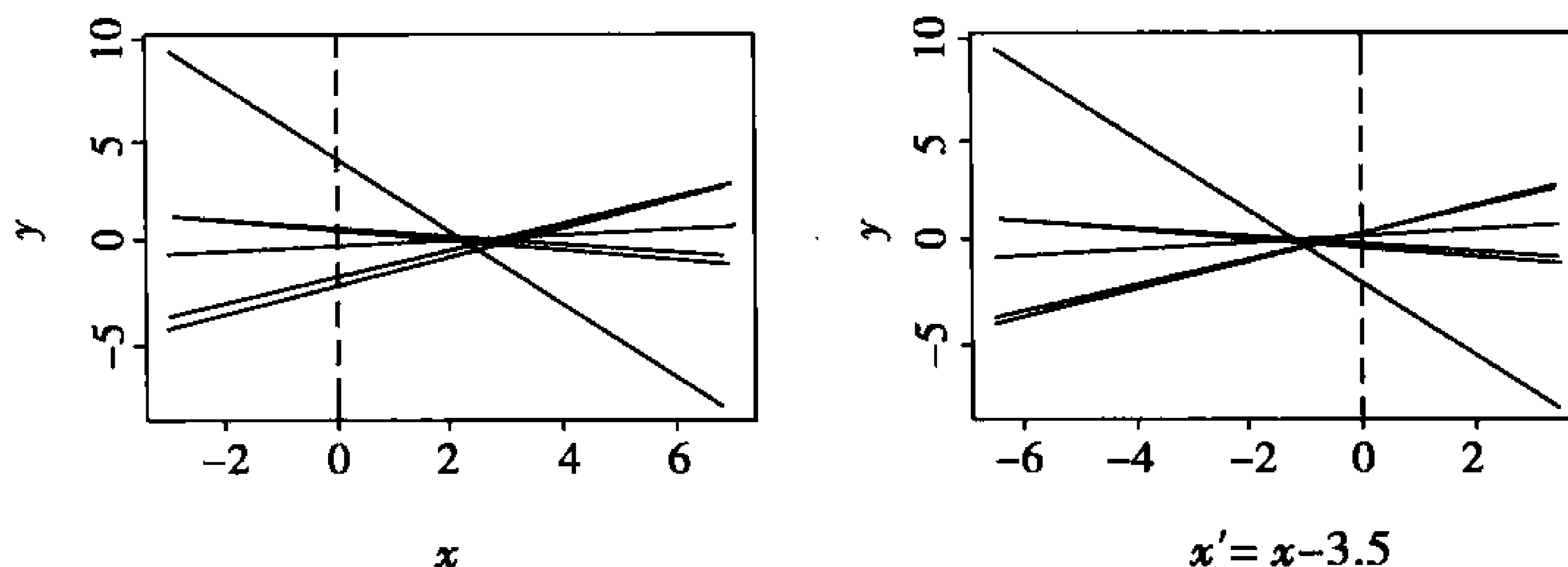



**图3.1 随机系数模型的特定群组的回归线,**
**说明解释变量转化后恒定性的缺乏**

用向量符号来表示随机系数模型的简化形式是很方便的,则

$$
\begin{aligned}
y_{ij} &= \underbrace{\gamma_{00}}_{\beta_0} + \underbrace{\gamma_{01}\omega_j}_{\beta_1 x_{1ij}} + \underbrace{\gamma_{10}x_{ij}}_{\beta_2 x_{2ij}} + \underbrace{\gamma_{11}(\omega_j x_{ij})}_{\beta_3 x_{3ij}} + \zeta_{0j} + \underbrace{\zeta_{1j}x_{ij}}_{\zeta_{1j}\, z_{1ij}} + \epsilon_{ij} \\
&= (\beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_3 x_{3ij}) + (\zeta_{0j} + \zeta_{1j}z_{1ij}) + \epsilon_{ij} \\
&= \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{\zeta}_j + \epsilon_{ij} \qquad (3.12)
\end{aligned}
$$

其中,$x_{qij}$是指带有固定系数的协变量,$z_{qij}$是带有随机系数的协变量。固定效应 $\boldsymbol{\beta}'=(\beta_0,\beta_1,\beta_2,\beta_3)$ 的协变量向量 $\mathbf{x}'_{ij}=(1,x_{1ij},x_{2ij},x_{3ij})$,包含了学生和学校层次的协变量,同时也包括学生和学校层次变量的乘积——代表着**跨层次交互作用**(cross-level interactions),$\mathbf{z}'_{ij}=(1,z_{1ij})$ 和 $\boldsymbol{\zeta}'_j=(\zeta_{0j},\zeta_{1j})$。

双层次模型可通过对应在式(3.9)和式(3.10)中的第一层次和第二层次模型来进行设定,或者可以直接通过式(3.12)中它们的简化形式来设定。前一种方法的运用参见 Raudenbush and Bryk(2002),后一种方法的范例参见 Goldstein(2003)、Rabe-Hesketh et al.(2004a)。

还可参见3.6.1节关于随机系数模型或纵贯性数据的"增长曲线模型"的讨论。

**双层次模型的矩阵符号**

通过将所有响应叠加到一个单一向量 $\mathbf{y}$ 上,一个双层次模型可以用矩阵符号的形式表示出来。相应的,我们将行向量 $\mathbf{x}'_{ij}$叠加到矩阵 $\mathbf{X}$ 中。然后我们将运用标准的术语,将这个矩阵表示为一个"设计(design)"矩阵,尽管我们承认它的值并不必然由一个实验设计所能决定(参见 Kempthorne,1980)。用 $\boldsymbol{\zeta}_{(D)}$ 指代所有随机效应的向量,整个样本的矩阵方程式变为

$$
\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_{(D)}\boldsymbol{\zeta}_{(D)} + \boldsymbol{\epsilon} \qquad (3.13)
$$

其中,下标(D)表示该矩阵或向量包含整个数据集的元素(这里 $\mathbf{y}\equiv\mathbf{y}_{(D)}$,$\mathbf{X}\equiv\mathbf{X}_{(D)}$ 且 $\boldsymbol{\epsilon}\equiv\boldsymbol{\epsilon}_{(D)}$)。注意 $\mathbf{Z}_{(D)}$ 是一个块对角矩阵,具有与第二层次的单元相一致的块。现在来考虑拥有两个协变量——它们同时具有固定系数和随机系数——的模型,其中,$x_{1ij}=z_{1ij}$,$x_{2ij}=z_{2ij}$。对于两个第二层次的单元,其中第一个包含3个第一层次的单元,第二个包含4个第一层次的单元,单个第一层次单元的矩阵方程详细地写在展示3.1中的B.i部分。

对于一个单一的第二层次单元 $j$,模型变为

$$
\mathbf{y}_{j(2)} = \mathbf{X}_{j(2)}\boldsymbol{\beta} + \mathbf{Z}_{j(2)}\boldsymbol{\zeta}_j + \boldsymbol{\epsilon}_{j(2)} \qquad (3.14)
$$

其中,$\mathbf{y}_{j(2)}$,$\mathbf{X}_{j(2)}$,$\boldsymbol{\zeta}_j$ 和 $\boldsymbol{\epsilon}_{j(2)}$是展示3.1中B.i部分单元层次表述的行,与第 $j$ 个第二层次单元相一致,但 $\mathbf{Z}_{j(2)}$是随机效应的块对角设计矩阵中相关的块,参见展示3.1中C.i部

分。这里下标 $j(2)$ 表示向量和矩阵包含第 $j$ 个第二层次单元的所有元素，注意 $\zeta_j=\zeta_{j(2)}$。

设 $\mathbf{\Psi}$ 为 $\zeta_j$ 的协方差矩阵，一个第二层次单元响应的条件协方差结构可写为

$$\mathbf{\Omega}_{j(2)} \equiv \mathrm{Cov}(y_{j(2)} \mid \mathbf{X}_{j(2)}, \mathbf{Z}_{j(2)}) = \mathbf{Z}_{j(2)}\mathbf{\Psi}\mathbf{Z}'_{j(2)} + \sigma^2_\epsilon \mathbf{I} \tag{3.15}$$

---

展示 3.1 一个双层次随机系数模型的矩阵符号形式

---

A. 单一单元的表示法：

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \zeta_{0j} + \zeta_{1j}z_{1ij} + \zeta_{2j}z_{2ij} + \epsilon_{ij}$$

B. 整个样本的表示法：

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_{(D)}\boldsymbol{\zeta}_{(D)} + \boldsymbol{\epsilon}$$

i. 单元层次表示（第一层次），针对响应 $y_{ij}$：

$$\underbrace{\begin{bmatrix} y_{11} \\ y_{21} \\ y_{31} \\ \hline y_{12} \\ y_{22} \\ y_{32} \\ y_{42} \end{bmatrix}}_{\mathbf{y}} = \underbrace{\begin{bmatrix} 1 & x_{111} & x_{211} \\ 1 & x_{121} & x_{221} \\ 1 & x_{131} & x_{231} \\ \hline 1 & x_{112} & x_{212} \\ 1 & x_{122} & x_{222} \\ 1 & x_{132} & x_{232} \\ 1 & x_{142} & x_{242} \end{bmatrix}}_{\mathbf{X}} \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}}_{\boldsymbol{\beta}} + \underbrace{\left[\begin{array}{ccc|ccc} 1 & z_{111} & z_{211} & 0 & 0 & 0 \\ 1 & z_{121} & z_{221} & 0 & 0 & 0 \\ 1 & z_{131} & z_{231} & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 & z_{112} & z_{212} \\ 0 & 0 & 0 & 1 & z_{122} & z_{222} \\ 0 & 0 & 0 & 1 & z_{132} & z_{232} \\ 0 & 0 & 0 & 1 & z_{142} & z_{242} \end{array}\right]}_{\mathbf{Z}_{(D)}} \underbrace{\begin{bmatrix} \zeta_{01} \\ \zeta_{11} \\ \zeta_{21} \\ \hline \zeta_{02} \\ \zeta_{12} \\ \zeta_{22} \end{bmatrix}}_{\boldsymbol{\zeta}_{(D)}} + \underbrace{\begin{bmatrix} \epsilon_{11} \\ \epsilon_{21} \\ \epsilon_{31} \\ \hline \epsilon_{12} \\ \epsilon_{22} \\ \epsilon_{32} \\ \epsilon_{42} \end{bmatrix}}_{\boldsymbol{\epsilon}}$$

ii. 群组层次表示（第二层次），针对响应 $\mathbf{y}_{j(2)}$：

$$\begin{bmatrix} \mathbf{y}_{1(2)} \\ \hline \mathbf{y}_{2(2)} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_{1(2)} \\ \hline \mathbf{X}_{2(2)} \end{bmatrix}\boldsymbol{\beta} + \left[\begin{array}{c|c} \mathbf{Z}_{1(2)} & 0 \\ \hline 0 & \mathbf{Z}_{2(2)} \end{array}\right] \begin{bmatrix} \boldsymbol{\zeta}_1 \\ \hline \boldsymbol{\zeta}_2 \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon}_{1(2)} \\ \hline \boldsymbol{\epsilon}_{2(2)} \end{bmatrix}$$

C. 单一群组的表示法：

$$\mathbf{y}_{j(2)} = \mathbf{X}_{j(2)}\boldsymbol{\beta} + \mathbf{Z}_{j(2)}\boldsymbol{\zeta}_j + \boldsymbol{\epsilon}_{j(2)}$$

i. 单元层次表示（第二层次），针对响应 $y_{ij}$（这里群组 $j=1$）：

$$\underbrace{\begin{bmatrix} y_{1j} \\ y_{2j} \\ y_{3j} \end{bmatrix}}_{\mathbf{y}_{j(2)}} = \underbrace{\begin{bmatrix} 1 & x_{11j} & x_{21j} \\ 1 & x_{12j} & x_{22j} \\ 1 & x_{13j} & x_{23j} \end{bmatrix}}_{\mathbf{X}_{j(2)}} \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}}_{\boldsymbol{\beta}} + \underbrace{\begin{bmatrix} 1 & z_{11j} & z_{21j} \\ 1 & z_{12j} & z_{22j} \\ 1 & z_{13j} & z_{23j} \end{bmatrix}}_{\mathbf{Z}_{j(2)}} \underbrace{\begin{bmatrix} \zeta_{0j} \\ \zeta_{1j} \\ \zeta_{2j} \end{bmatrix}}_{\boldsymbol{\zeta}_j} + \underbrace{\begin{bmatrix} \epsilon_{1j} \\ \epsilon_{2j} \\ \epsilon_{3j} \end{bmatrix}}_{\boldsymbol{\epsilon}_{j(2)}}$$

---

### 3.2.3 三层次模型

既然学生也是嵌套于班级之中（而班级又嵌套于学校中），那么一个针对嵌套于学校当中的学生的双层次模型就可能是不现实的。我们会因此期望同一所学校中属于同一个班级的两个学生的响应之间具有更高的相关性。这可以运用一个三层次模型来进行建模。

扩展式(3.12)，一个关于第一层次单元 $i$（如学生）、第二层次单元 $j$（如班级）和第三层次单元 $k$（如学校）的广义三层次模型，可写为如下简化形式

$$y_{ijk} = \mathbf{x}'_{ijk}\boldsymbol{\beta} + \mathbf{z}^{(2)\prime}_{ijk}\boldsymbol{\zeta}^{(2)}_{jk} + \mathbf{z}^{(3)\prime}_{ijk}\boldsymbol{\zeta}^{(3)}_{k} + \epsilon_{ijk} \tag{3.16}$$

这些项分别代表模型的固定部分、第二层次随机部分、第三层次随机部分和第一层次的残差。$\mathbf{x}_{ijk}$是解释变量(包括常数)的一个向量,带有固定回归系数$\boldsymbol{\beta}$,$\mathbf{z}_{ijk}^{(2)}$是解释变量的一个$M_2$维向量,带有第二层次的随机系数$\boldsymbol{\zeta}_{jk}^{(2)}$,$\mathbf{z}_{ijk}^{(3)}$是解释变量的一个$M_3$维向量,带有第三层次的随机系数$\boldsymbol{\zeta}_k^{(3)}$。随机效应上的上标表示的是它们变动的层次,而协变量上的上标表示的是相应的随机系数变动的层次。每个层次上的随机效应都具有一个多元正态分布,并且不同层次上的随机效应是相互独立的,且独立于第一层次的残差。

扩展式(3.13),整个样本的三层次模型可用矩阵符号形式写为

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_{(D)}^{(2)}\boldsymbol{\zeta}_{(D)}^{(2)} + \mathbf{Z}_{(D)}^{(3)}\boldsymbol{\zeta}_{(D)}^{(3)} + \boldsymbol{\epsilon} \tag{3.17}$$

有时,运用下述这样一个单一的设计矩阵是非常便利的,即

$$\mathbf{Z}_{(D)} \equiv [\mathbf{Z}_{(D)}^{(2)}, \mathbf{Z}_{(D)}^{(3)}]$$

带有一个相应的随机效应的向量

$$\boldsymbol{\zeta}_{(D)} \equiv [\boldsymbol{\zeta}_{(D)}^{(2)\prime}, \boldsymbol{\zeta}_{(D)}^{(3)\prime}]'$$

其中,$\boldsymbol{\zeta}_{(D)}$代表整个数据集的所有随机效应,同时包括第二层次和第三层次的随机效应。于是上述三层次模型可以表达为

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_{(D)}\boldsymbol{\zeta}_{(D)} + \boldsymbol{\epsilon} \tag{3.18}$$

该方程式在关于随机截距模型的展示3.2中B.i部分有详细描述。这里,两个第三层次单元,每一个都包含有两个第二层次的单元,这两个第二层次单元又都包含有两个第一层次的单元。我们已经改变$\mathbf{Z}_{(D)}$中列的顺序,以便获得一个块对角的形式,含有对应第一个第三层次群组中单元的块$\mathbf{Z}_{1(3)}$和对应第二个第三层次群组中单元的块$\mathbf{Z}_{2(3)}$。随机效应$\boldsymbol{\zeta}_{k(3)}$的向量也相应地改变了顺序。这就使第$k$个第三层次单元的模型可写为

$$\mathbf{y}_{k(3)} = \mathbf{X}_{k(3)}\boldsymbol{\beta} + \mathbf{Z}_{k(3)}\boldsymbol{\zeta}_{k(3)} + \boldsymbol{\epsilon}_{k(3)}$$

参见展示3.2中的C.i部分。

现在模型看起来代数学地等价于一个双层次模型。这有一个优势,即我们可将任何针对双层次模型特别发展出的结果应用于更高层次的模型当中。例如,我们可以直接应用方程式(3.15)来获得第$k$个第三层次单元响应的条件协方差矩阵

$$\boldsymbol{\Omega}_{k(3)} \equiv \mathrm{Cov}(\mathbf{y}_{k(3)} \mid \mathbf{X}_{k(3)}, \mathbf{Z}_{k(3)}) = \mathbf{Z}_{k(3)}\boldsymbol{\Psi}_{k(3)}\mathbf{Z}'_{k(3)} + \sigma_\epsilon^2\mathbf{I} \tag{3.19}$$

其中,$\boldsymbol{\Psi}_{k(3)}$是$\zeta_{k(3)}$的协方差矩阵,也即第$k$个第三层次单元的所有随机效应。

### 3.2.4 更高层次的模型

一个广义的$L$层次模型可写为

$$y = \mathbf{x}'\boldsymbol{\beta} + \sum_{l=2}^{L}\mathbf{z}^{(l)\prime}\boldsymbol{\zeta}^{(l)} + \boldsymbol{\epsilon} \tag{3.20}$$

其中,固定部分在前,$\mathbf{Z}^{(l)}$是解释变量的一个$M_l$维向量,含有第$l$层次上的随机系数$\boldsymbol{\zeta}^{(l)}$,并且我们已经去掉单元和群组的指数来简化符号形式。在一个给定的第$l$层次上的随机效应,通常被假定为具有一个带零均值和协方差矩阵$\boldsymbol{\Psi}^{(l)}$的多元正态分布。不同层次上的随机效应则被假定为相互独立的,且独立于残差项(residual error term)。

### 3.2.5 广义线性混合模型

到目前为止,我们已经讨论了连续响应的多层次模型。然而,所有第2章中讨论的响

## 展示 3.2　一个三层次随机截距模型的矩阵符号形式

A. 单一单元的表示法：

$$y_{ijk} = \mathbf{x}'_{ijk}\boldsymbol{\beta} + \zeta^{(2)}_{jk} + \zeta^{(3)}_{k} + \epsilon_{ijk}$$

B. 整个样本的表示法：

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}^{(2)}_{(D)}\boldsymbol{\zeta}^{(2)}_{(D)} + \mathbf{Z}^{(3)}_{(D)}\boldsymbol{\zeta}^{(3)}_{(D)} + \boldsymbol{\epsilon}$$
$$= \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_{(D)}\boldsymbol{\zeta}_{(D)} + \boldsymbol{\epsilon}$$

i. 单元层次表示法（第一层次），针对响应 $y_{ijk}$：

$$\underbrace{\begin{bmatrix} y_{111} \\ y_{211} \\ y_{121} \\ y_{221} \\ \hline y_{112} \\ y_{212} \\ y_{122} \\ y_{222} \end{bmatrix}}_{y} = \mathbf{X}\boldsymbol{\beta} + \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\mathbf{Z}^{(2)}_{(D)}} \underbrace{\begin{bmatrix} \zeta^{(2)}_{11} \\ \zeta^{(2)}_{21} \\ \hline \zeta^{(2)}_{12} \\ \zeta^{(2)}_{22} \end{bmatrix}}_{\boldsymbol{\zeta}^{(2)}_{(D)}} + \underbrace{\begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ \hline 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}}_{\mathbf{Z}^{(3)}_{(D)}} \underbrace{\begin{bmatrix} \zeta^{(3)}_{1} \\ \hline \zeta^{(3)}_{2} \end{bmatrix}}_{\boldsymbol{\zeta}^{(3)}_{(D)}} + \boldsymbol{\epsilon}$$

$$= \mathbf{X}\boldsymbol{\beta} + \underbrace{\left[\begin{array}{ccc|ccc} 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{array}\right]}_{\mathbf{Z}_{(D)}} \underbrace{\begin{bmatrix} \zeta^{(2)}_{11} \\ \zeta^{(2)}_{21} \\ \zeta^{(3)}_{1} \\ \hline \zeta^{(2)}_{12} \\ \zeta^{(2)}_{22} \\ \zeta^{(3)}_{2} \end{bmatrix}}_{\boldsymbol{\zeta}_{(D)}} + \boldsymbol{\epsilon}$$

ii. 群组层次表示法（第三层次），针对响应 $\mathbf{y}_{k(3)}$：

$$\begin{bmatrix} \mathbf{y}_{1(3)} \\ \hline \mathbf{y}_{2(3)} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_{1(3)} \\ \hline \mathbf{X}_{2(3)} \end{bmatrix}\boldsymbol{\beta} + \left[\begin{array}{c|c} \mathbf{Z}_{1(3)} & 0 \\ \hline 0 & \mathbf{Z}_{2(3)} \end{array}\right] \begin{bmatrix} \boldsymbol{\zeta}_{1(3)} \\ \hline \boldsymbol{\zeta}_{2(3)} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon}_{1(2)} \\ \hline \boldsymbol{\epsilon}_{2(2)} \end{bmatrix}$$

C. 高层次群组 $k$ 的表示法：

$$\mathbf{y}_{k(3)} = \mathbf{X}_{k(3)}\boldsymbol{\beta} + \mathbf{Z}_{k(3)}\boldsymbol{\zeta}_{k(3)} + \boldsymbol{\epsilon}_{k(3)}$$

i. 单元层次表示法（第一层次），针对响应 $y_{ijk}$：

$$\underbrace{\begin{bmatrix} y_{11k} \\ y_{21k} \\ y_{12k} \\ y_{22k} \end{bmatrix}}_{\mathbf{y}_{k(3)}} = \mathbf{X}_{k(3)}\boldsymbol{\beta} + \underbrace{\begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}}_{\mathbf{Z}_{k(3)}} \underbrace{\begin{bmatrix} \zeta^{(2)}_{1k} \\ \zeta^{(2)}_{2k} \\ \zeta^{(3)}_{k} \end{bmatrix}}_{\boldsymbol{\zeta}_{k(3)}} + \underbrace{\begin{bmatrix} \epsilon_{11k} \\ \epsilon_{21k} \\ \epsilon_{12k} \\ \epsilon_{22k} \end{bmatrix}}_{\boldsymbol{\epsilon}_{k(3)}}$$

应类型，都能通过设定响应的条件分布而适用模型，已知随机效应为一个带有线性估计量 $\nu$ 的广义线性模型，线性估计量 $\nu$ 与式(3.20)的条件均值具有同一形式，即

$$\nu = \mathbf{z}'\boldsymbol{\beta} + \sum_{l=2}^{L} \mathbf{z}^{(l)'}\boldsymbol{\zeta}^{(l)}$$

这个结果模型被称为一个**广义线性混合模型**。线性混合模型是一个同一性链接和服从条件正态分布的响应的特殊情况。

在广义线性混合模型中，回归系数代表协变量的**条件效应**，给出了随机效应的数值。这些效应可被理解为特定群组的效应。相反，边际的或**总体平均**的效应，是整合了随机效应之后的协变量的效应。对于一个随机截距正态概率模型，条件效应和边际效应的区别如1.4 节的图1.6 所示。注意，这里的边际效应概念与计量经济学中的某些用法不同，在计量经济学中，它典型地是指在没有潜变量的模型中，一个协变量（已知其他的协变量）中的微小变化对期望响应的效应（如 Greene，2003）。对于一个同一性链接，条件效应等于边际效应，但这不是普遍现象。我们将在3.6.5 节进行深入讨论，在4.8.1 节讨论其推导过程，在9.2 节给出一个实例。

在广义线性模型中，响应的方差（已知随机效应、协变量和"第一层次方差"）取决于具有设定条件分布的方差函数。如果这些响应是计数，被建模为条件泊松或二项分布，那么第一层次上的过度离散则可通过纳入一个第一层次上的随机截距而进行模型建构（如 Breslow and Clayton，1993）。戈德斯坦恩（Goldstein，1987）和夏尔（Schall，1991）以及其他学者采用了一种拟似然方法，在第一层次的方差函数中纳入一个额外的离散参数（参见2.3.1 节）。

既然第一层次的方差通常不是固定的，那么在同一个群组中，观测到的响应之间的相关性也不是固定的，即便是在一个简便的随机截距模型中（如 Goldstein et al.，2002）。对于二项响应或定序响应，组内相关性通常是根据潜响应 $y^*$ 之间的相关性（它是固定的）来进行表达。对于一个随机截距正态概率模型，这种类型的组内相关性即为

$$\rho \equiv \mathrm{Cor}(y_{ij}^*, y_{i'j}^* \mid x_{ij}, x_{i'j}) = \frac{\psi}{\psi + 1}$$

即众所周知的二项情形中没有协变量的**四项相关**（tetrachoric correlation）。对于一个对数概率模型，分母中的"1"被替换为 $\pi^2/3$，即第一层次对数概率误差的方差。

### 3.2.6 未分层随机效应的模型

**交叉随机效应模型**

迄今为止，我们已讨论了层级模型，其中单元通过某些因子（如学校）被分类进高层次群组第 $L$ 层次中。每一个高层次群组中的单元，然后通过一个另外的因子（如班级）被（低）分进一个 $L-1$ 层次的群组中，等等。定义分类的因子都是在这个意义上嵌套的，即一个低层次的群组只能隶属于一个更高层次的群组（如一个班级只能属于一个学校）。

我们现在讨论未分层次的模型，其中单元是通过两个或更多因子而**交叉分类**的（cross-classified），每个单元又潜在地属于不同因子中各个层次的任意组合。一个突出的例子是面板数据，其中，因子"个人"（或国家、公司等）是与另一个因子"时间"或"时机"交叉在一起的。虽然特定单元未观测到的异质性经常运用随机效应来进行调节（参见3.6.1节），但是特定时刻未观测到的异质性的随机效应建模，应归功于各个时刻事件的

共享经验,如罢工、新的立法或天气状况,这个共享的经验似乎是被局限于计量经济学中。如果两方面因子都被认为是随机的,那么计量经济学家将这个模型称为一个双向误差成分模型(two-way error component model)(如 Baltagi,2001,第 3 章)。交叉分类的随机效应模型也出现在"概化理论(generalizability theory)"中。这里,一个简便的设计是一个关于评价对象和评分人的双向交叉分类,可参见 3.3.1 节。

现在来考虑用小学和中学进行交叉分类的学生。忽略小学,针对嵌套在中学 $j$ 中的学生 $i$ 的一个双层次随机截距模型可被设定为

$$\nu_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \zeta_j^{(2)} \tag{3.21}$$

嵌套在中学当中的学生的相应层级结构如图 3.2 所示,其中每个学生是由一个垂直的短线来表示,右边垂直的长线代表他或她的中学。

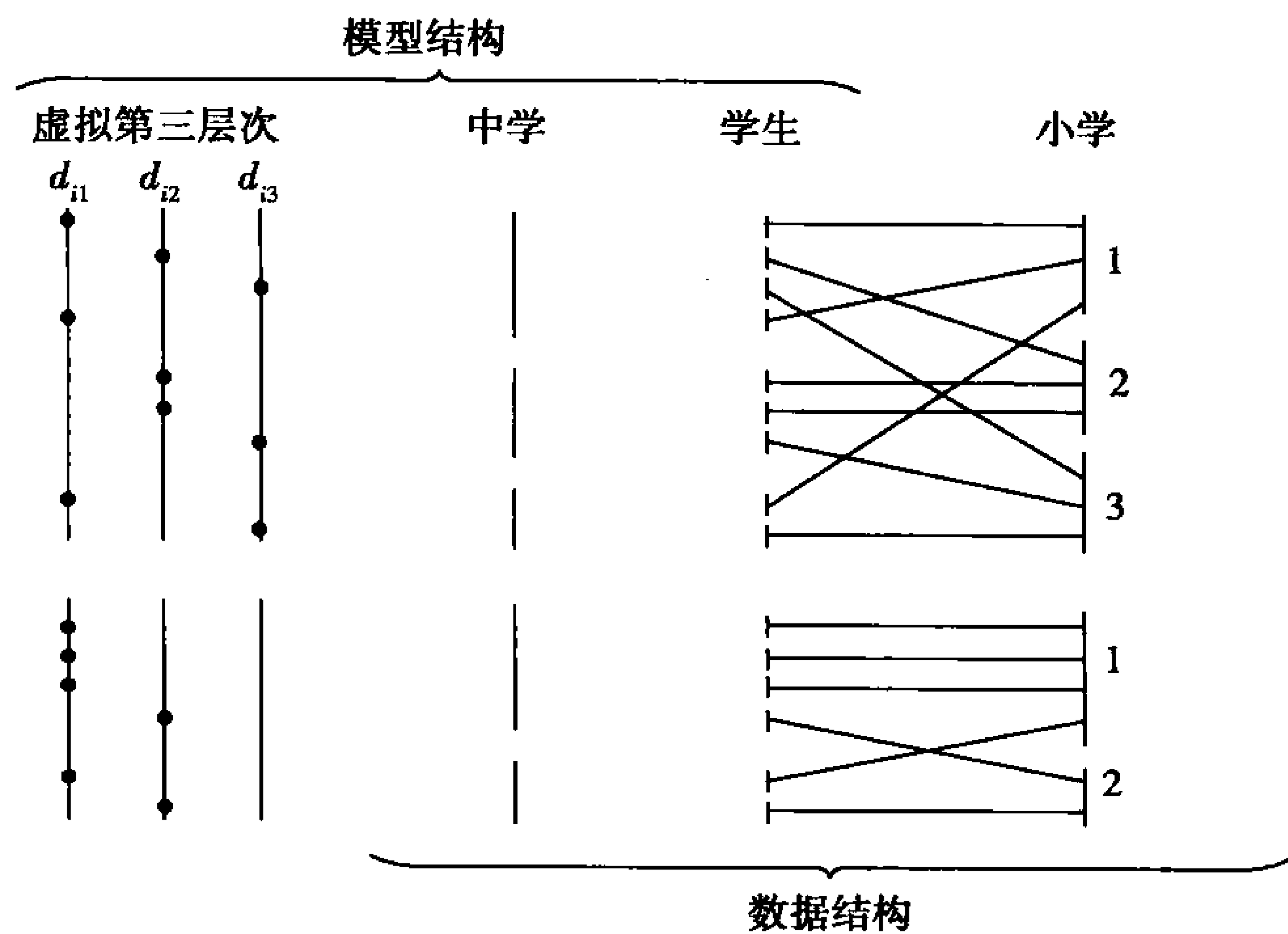



**图 3.2 小学有交叉的中学学生的模型结构和数据结构**

为学生的小学 $p$ 纳入一个额外的随机截距 $\zeta_p$,就破坏了这种嵌套,因为来自同一所中学的学生未必来自于同一所小学,反过来也这样。在图 3.2 中可知,连接学生与其小学的线条是交叉的。注意,不能改变学生的组别来解开这些交叉。交叉随机效应模型可写为

$$\nu_{ijp} = \mathbf{x}'_{ijp}\boldsymbol{\beta} + \zeta_j^{(2)} + \zeta_p$$

其中,下标 $p$ 已经被增加到固定部分中,以便调节特定小学协变量。

戈德斯坦恩(Goldstein,1987)提出了一个诀窍,将一个交叉随机效应模型表示为一个带有大量随机效应的层级模型。这是非常重要的,因为很多估计方法都被局限于嵌套的随机效应模型中。在当前的例子里,我们首先必须引入一个"虚拟"层次,小学和中学都嵌套于这个层次中,如城镇,我们称其为第三层次。注意这个层次在任何意义上都不需要是"真实的";例如,如果一个孩子搬到另一个城镇去读中学,那么这两个城镇可以合并为一个单一的第三层次单元。如果不可能找到一个虚拟的第三层次,那么这个层次可以被定义为一个包含全部数据集的单一单元。在图 3.2 中,虚拟层次是由中学和小学生成的组别垂直线所表示。在一个虚拟第三层次单元中的学生,只能隶属于该单元中的中学和小学(不能在第三层次单元之间有交叉)。

现在,在每一个第三层次单元内,任意地将小学标记为 $p' = 1, \cdots, n_{\max}$ 其中,$n_{\max}$ 是一个第三层次单元中小学的最大数。对第三层次使用下标 $k$,我们可以将上面的模型等价地写为

$$\nu_{ijp} = \mathbf{x}'_{ijp}\boldsymbol{\beta} + \zeta_j^{(2)} + \sum_{p'=1}^{n_{\max}} \zeta_{kp'}^{(3)} d_{ip'}$$

其中,$\zeta_{kp'}^{(3)}$是第三层次上的第 $p'$个随机"斜率";当学生 $i$ 进入任何一个标记为 $p'$的小学时,$d_{ip'}$等于1,否则为0。在此,所有编号为 $p'$的小学的随机截距,都是由(有不同的实现方式)一个小学 $p'$的虚拟变量的随机斜率(在第三层次上是变动的)来表示。这些随机斜率的方差都被约束为是相等的(目的是在原始模型中获得一个固定的随机截距方差),并且它们的协方差被设置为0(因为小学都是相互独立的)。中学的随机斜率和随机截距之间的协方差也是0。

在 $n_{\max}=3$ 的图中,一个虚拟第三层次单元中的小学,是被标记为从1到最多是3。虚拟第三层次的3条长垂直线,代表被标记为1,2和3的小学的随机效应值。圆点表示的是,对于那些学生,相应的虚拟变量 $d_{i1}$到 $d_{i3}$都等于1。例如,顶端的前3个圆点,表示顶端的前3个学生分别属于小学1、小学2和小学3。一个单一随机效应可运用于几所小学(所有小学都有相同标记)的原因,是因为每一所小学的随机效应都有不同的值。

遗憾的是,当 $n_{\max}$变得太大时,用交叉效应将模型表述为多层次模型则行不通。在这种情况下,马尔可夫链蒙特卡洛方法(如6.11.5节讨论的AIP算法)则非常有用。

关于交叉随机效应模型的进一步讨论可参见 Snijders and Bosker(1999,第11章)、Raudenbush and Bryk(2002,第12章)、Goldstein(2003,附录11.1)。

**多隶属模型**(Multiple-membership model)

忽略小学,我们现在来看式(3.21)中含有一个针对中学的随机截距的模型。如果一个学生 $i$ 已经上过几所中学$\{j\}$,在学校 $h\in\{j\}$中学习的时间为 $t_{ih}$,那么一个合理的"多隶属"模型则为

$$\nu_{i\{j\}} = \mathbf{x}'_i\boldsymbol{\beta} + \sum_{h\in\{j\}} \zeta_h^{(2)} t_{ih}$$

其中,$\zeta_h^{(2)}$ 代表进入学校 $h$ 的效应(每次单元的"显露"时间)(有时,通过将总时间 $\sum_h t_{ih}$ 进行划分,$t_{ih}$被替代为在学校 $h$ 中所花费的时间比例)。

简便起见,我们假设不存在特定学校的固定效应。有两个问题导致这个模型是非嵌套的。第一,对于上过(至少)两所学校的学生来说,进过的第一所和第二所学校是交叉的,即进入的第一所学校是同一个的那些学生,可能会毕业于不同的第二所学校,反之亦然。第二,一个特定学校的随机效应必须取值相同,不论它是一个已知学生的第一所、第二所或第三所学校。这一方面是与交叉效应模型不同的。

希尔和戈德斯坦恩(Hill and Goldstein,1998)提出了一种方法,将这些模型表示为层级模型,也可参考 Rasbash and Browne(2001)。就本例而言,我们首先需要找到一个第三层次,这样一来,学生则仅仅在一个层次内在学校之间交叉,而不会在不同层次间交叉(如城镇)。将每一个第三层次单元中的学校任意地标记为 $h=1,\cdots,n_{\max}$,其中,$n_{\max}$是每个第三层次单元的最大学校数,模型可写为

$$\nu_{i\{j\}} = \mathbf{x}'_i\boldsymbol{\beta} + \sum_{h=1}^{n_{\max}} \zeta_{kh}^{(3)} t_{ih} d_{ih}$$

其中,$\zeta_{kh}^{(3)}$是第三层次上的第 $h$ 个随机斜率,$d_{ih}$是一个虚拟变量,如果学生 $i$ 曾经进入了任何一个标记为 $h$ 的学校,则 $d_{ih}$等于1,否则为0。

## 3.3 因子模型和题器响应模型

### 3.3.1 理想的真实得分:测量模型

理想的真实得分是存在的,原则上可以被测量,这样也就可以代表一种**黄金标准**(gold standard)。然而,由于现实原因,人们经常采用易于出错的测量而不是黄金标准。例如,罗德尔等人(Roeder et al.,1996)提到的一项应用,其中,低密度脂蛋白胆固醇的测量是通过运用低成本和低时间消耗的"总胆固醇"测量而得到的。

一个简便测量模型——在"经典检验理论"中是标准的(参见 Lord and Novick, 1968)——可被写为

$$y_{ij} = \eta_j + e_{ij} \tag{3.22}$$

其中,$y_{ij}$是在第$j$个单元上的第$i$个测量,$\eta_j$是单元$j$的真实得分,含有方差$\psi$,而$e_{ij}$是含有方差$\theta$的测量误差。

通常我们假设$e_{ij}$是相互独立的,这样$y_{ij}$就条件独立于给定的$\eta_j$。此外,还假设$e_{ij}$具有零期望并且独立于真实得分。然而,如果有一个校验样本,对于它来说,除了易错的测量之外,黄金标准是现实可用的,那么我们就可以评估这些假设。特别指出的是,如果测量误差的期望值不是零,那么这个测量就是**有偏的**。

信度$\rho$可以被定义为测量的真实得分方差在总方差中的比例,即

$$\rho = \frac{\mathrm{Var}(\eta_j)}{\mathrm{Var}(\eta_j) + \mathrm{Var}(e_{ij})} = \frac{\psi}{\psi + \theta}$$

注意这正好是一个随机截距或**单向随机效应**(one-way random effects)模型的组内相关,参见3.2.1节的讨论。

如果对每个人的测量可以被看成是可互换的重复,那么简便的单向随机效应模型就是适当的,但是这种方式是不现实的。例如,如果考试是通过一个由主考官或评分人$i$构成的面板来标示的,那么一个双向模型

$$y_{ij} = \beta_i + \eta_j + e_{ij}$$

可能更加适合,其中,$\beta_i$表示的是评分人$i$的**响应偏差**。

如果这些评分人是可能的评分人中的一个随机样本,那么$\beta_i$则可被视为随机的,得到一个**双向随机效应**模型。注意,如果每个应试者都被每一个评分人评价,那么评分对象和评分人的随机效应就不是嵌套的,而是交叉的(参见3.2.6节)。

如果这些评分人是固定的,那么$\beta_i$就是固定效应,并且这个模型是一个**双向混合效应模型**。如果每个评分人对每个对象提供同样的测量,那么对象-评分人的交互作用则可被包括在这两种类型的模型中。这些模型以及可以从它们推导出的不同类型的信度系数,在夏诺特和弗雷斯(Shrout and Fleiss,1979)和麦克劳和黄(McGraw and Wong,1996)的著作中有所讨论。

将$\beta_i$看做是固定的,我们可以允许测量尺度和信度在不同方法之间(如评分人)是不同的。这种**同属测量模型**(congeneric measurement model)(Jöreskog,1971b)可以被设定为

$$y_{ij} = \beta_i + \lambda_i\eta_j + e_{ij} \tag{3.23}$$

其中,$\beta_i$是固定参数,$\mathrm{E}(\eta_j)=0$,$\psi \equiv \mathrm{Var}(\eta_j)$,$\theta_{ii} \equiv \mathrm{Var}(e_{ij})$。在这个模型中,$\beta_i$依旧表示

评分人的偏差，而 $\lambda_i$ 是一个特定评分人尺度参数。我们可将 $\lambda_i\eta_j$ 理解为在评分人 $i$ 的单元中测量的真实得分。

通过设 $\lambda_1=1$，通常真实得分 $\eta_j$ 的尺度被固定到 $y_{1j}$的尺度中，该类型的一种辨识约束在5.2.3节进行了讨论。测量误差方差 $\theta_{ii}$在不同的评分人之间可能有所不同。评分人 $i$ 的信度变为

$$\rho_i=\frac{\mathrm{Var}(\lambda_i\eta_j)}{\mathrm{Var}(\lambda_i\eta_j)+\mathrm{Var}(e_{ij})}=\frac{\lambda_i^2\psi}{\lambda_i^2\psi+\theta_{ii}}$$

同属测量模型包含了下述针对单元 $j$ 的测量 $y_j$ 的向量的协方差结构，即

$$\Omega\equiv\mathrm{Cov}(\mathbf{y}_j)=\begin{bmatrix}\lambda_1^2\psi+\theta_{11} & & & \\ \lambda_2\psi\lambda_1 & \lambda_2^2\psi+\theta_{22} & & \\ \vdots & \vdots & \ddots & \\ \lambda_I\psi\lambda_1 & \lambda_I\psi\lambda_2 & \cdots & \lambda_I^2\psi+\theta_{II}\end{bmatrix}$$

同属测量模型的3种常见的特殊情况如下：

- 基本Tau等价测量模型，其中，$\lambda_i=\lambda$。
- Tau等价测量模型，其中，$\lambda_i=\lambda$ 且 $\beta_i=\beta$。
- 平行测量模型，其中，$\lambda_i=\lambda$，$\beta_i=\beta$ 且 $\theta_{ii}=\theta$，得到式(3.22)中的简便测量模型。

将 $e_{ij}$解释为测量误差，预示着评分人在测量同一个事物。然而，如果评分人受到不同特质的影响，如笔迹或拼写，那么对评分人 $i$ 的期望则可表达为 $\beta_i+\lambda_i\eta_j+s_{ij}$，其中，$s_{ij}$ 是特定对象的偏差（如是由于笔迹或拼写造成的），或者换句话说，是一个评分人与对象的交互作用（如 Duun,1992）。如上文所提到的，特定因子的方差不能从测量误差方差中分离出来，除非对于每一个评分人和对象来说，都能使用重复的相同的测量。

将**伯克森测量模型**（Berkson,1950）从上文讨论过的经典测量模型中区分出来是至关重要的。在伯克森模型中，真实得分 $\eta_{ij}$是假定为环绕测量得分 $y_j$ 而正态分布的，即

$$\eta_{ij}=y_j+\epsilon_{Bij}$$

其中，$\epsilon_{Bij}$是**伯克森误差**。当 $y_j$ 是控制变量时，这一模型可能是适当的。例如，一名实验者可能打算执行药物 $y_j$ 的一个给定剂量，但是在 $i$ 时刻给出的实际真实剂量 $\eta_{ij}$，往往由于测量误差的存在而与 $y_j$ 不同。伯克森测量模型因此假设测量的响应是独立于测量误差的，但是经典测量模型并非如此。这对于3.5节讨论的包含协变量测量误差的回归具有重要的意义。

### 3.3.2 经典真实得分：公共因子模型

与理想真实得分不同，经典真实得分或假设构念即使在原则上也不能被直接测量，智力就是一个典型的例子。构念由不同的指标或题器来进行测量，如一个智力测试中的问题。与一个测量模型相反，从期望值可以被理解为“真实”智力的角度来说，单个题器不能用来测量智力本身。反而是智力的不同方面可以通过不同的题器进行测量；例如题器通常会需要不同的关于口头的、数量的、抽象的和直观的推理的“混合（blends）”。因此，对于一个特定题器的回答，既是对一般性智力的反映，也是对一个特定题器内容的反映，分别被称为公共因子和特殊因子（specific factor）。可以参考1.3节关于假设构念的更多讨论。

对于题器 $i=1,\cdots,I$，一维公共因子模型可写为

$$y_{ij} = \beta_i + \lambda_i \eta_j + \epsilon_{ij} \tag{3.24}$$

其中，$\eta_j$ 是对象 $j$ 的公共因子或潜在特征，$\lambda_i$ 是第 $i$ 个题器的一个因子负荷，$\epsilon_{ij}$是独特因子(unique factor)。我们定义 $\psi \equiv \text{Var}(\eta_j)$，$\theta_{ii} \equiv \text{Var}(\epsilon_{ij})$，并且令 $\eta_j$ 独立于 $\epsilon_{ij}$。

公共因子 $\eta_j$ 的尺度可通过"锚定"(通常是约束第一个因子负荷，$\lambda_1 = 1$)，或者因子标准化(factor standardization)(将因子方差固定为一个正的常数，$\psi = 1$)来进行固定。尽管从任一辨识约束得到的这些模型都是等价的(参见 5.3 节)，但是从"因子恒等性(factorial invariance)"观点的角度出发，锚定是非常有用的(参见 Meredith，1964，1993；Skrondal，1999)。例如，假定模型(3.24)是针对一个总体的，但是我们认为单元的子总体(subpopulation)具有负的因子值。这样，原始因子负荷就是从锚定(带有一个简化的方差估计 $\psi$)后的——而不是因子标准化后的——子总体中重新获得的。

注意，在这里的上下文中，截距 $\beta_i$ 不能被理解为测量偏差。如果特定题器均值 $\bar{y}_i$ 已经从 $y_{ij}$中减去，那么截距是可以省略掉的。还要注意，一维公共因子模型和式(3.23)中呈现的同属测量模型，是算术等价的。

独特因子可以进一步被分解为

$$\epsilon_{ij} = s_{ij} + e_{ij}$$

一个特殊因子的和与测量误差是相互独立的，并且独立于 $\eta_j$。注意，这个特殊因子有一个与测量模型中评分人和对象的交互作用相似的解释。特殊因子一般被认为是一个题器真实得分的一部分，其中信度变为

$$\rho_i = \frac{\lambda_i^2 \psi + \text{Var}(s_{ij})}{\lambda_i^2 \psi + \text{Var}(s_{ij}) + \text{Var}(e_{ij})}$$

遗憾的是，在大多数设计中，特殊因子和测量误差的方差不能被分别辨识出来，因为对于单个题器来说不存在复制品。根据纵贯性的或多方法-多特质设计(multimethod-multitrait design)，复制有时被应用于一种将独特因子分解成特殊因子和测量误差的尝试中(如 Alwin，1989 及其中所列的参考书目)。在没有复制的情况下，因子模型的信度经常有点随意地被表示为

$$\rho_i = \frac{\lambda_i^2 \psi}{\lambda_i^2 \psi + \text{Var}(s_{ij}) + \text{Var}(e_{ij})}$$

它表示了真实信度的较低的范围。

与这个信度紧密联系的是 Cronbach $\alpha$ 系数，它可以被理解为一个未加权总得分的信度的最大似然估计式(在没有复制的情况下进行估计)，通过假定题器是平行测量(如 Novick and Lewis，1967)。可以参考格里尼和卡迈因斯(Greene and Carmines，1980)关于总得分的 $\alpha$ 和其他信度测量的讨论。

因子模型也可以写成如下矩阵符号形式

$$\mathbf{y}_j = \boldsymbol{\beta} + \boldsymbol{\Lambda}\eta_j + \boldsymbol{\epsilon}_j$$

其中，$\boldsymbol{\beta}$ 是截距的一个 $I \times 1$ 向量，$\boldsymbol{\Lambda}$ 是因子负荷的一个 $I \times 1$ 向量，$\boldsymbol{\epsilon}_j$ 是特殊因子的一个 $I \times 1$ 向量，$I$ 是题器的总数。

协方差结构(在这里也称为一个**因子结构**)变为

$$\boldsymbol{\Omega} \equiv \text{Cov}(\mathbf{y}_j) = \boldsymbol{\Lambda}\psi\boldsymbol{\Lambda}' + \boldsymbol{\Theta} = \begin{bmatrix} \lambda_1^2\psi + \theta_{11} & & & \\ \lambda_2\psi\lambda_1 & \lambda_2^2\psi + \theta_{22} & & \\ \vdots & \vdots & \ddots & \\ \lambda_I\psi\lambda_1 & \lambda_I\psi\lambda_2 & \cdots & \lambda_I^2\psi + \theta_{II} \end{bmatrix}$$

其中，$\Theta$ 是一个对角矩阵，含有放在对角线上的 $\theta_{ii}$。注意，一维公共因子模型和同属测量模型的协方差结构是等价的。为了固定因子尺度，我们通常将一个因子负荷固定为 1，或者将因子方差固定为 1（参见 5.2.3 节）。

本节所讨论的模型都已经用题器进行了**反映**，被解释为反映的或者由一个潜变量所"造成"的。然而，有时将潜变量解释为**格式化的**（formative）——由题器所"造成"的——更有道理。一个权威的案例是关于一个家庭的社会经济地位（SES）测量，它以家庭中成年人的教育和收入为基础。在这个案例中，一个反映的模型是不确定的；我们会期望教育和收入影响 SES，而不是某些其他方式。在这种格式化的情况下，运用因子模型会引起一种设定错误。我们可以参考爱德沃兹和巴格兹（Edwards and Bagozzi，2000）关于题器与构念之间不同关系类型的概述。

### 3.3.3 多维因子模型

一维因子模型将一个相当限制性的结构强加到协方差上。在构造 $I(I+1)/2$ 个方差和协方差时，仅仅会用到 $2I$ 个参数。因此，低限制的多维因子模型通常是有用的。一个 $M$ 维因子模型可以用公式表达为

$$\begin{array}{ccccccccc} y_{1j} & = \beta_1 & + \lambda_{11}\eta_{1j} & + \cdots & + \lambda_{1M}\eta_{Mj} & + \epsilon_{1j} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ y_{Ij} & = \beta_I & + \lambda_{I1}\eta_{1j} & + \cdots & + \lambda_{IM}\eta_{Mj} & + \epsilon_{Ij} \end{array} \tag{3.25}$$

这个模型也可用矩阵形式表示为

$$\mathbf{y}_j = \boldsymbol{\beta} + \boldsymbol{\Lambda}_y\boldsymbol{\eta}_j + \boldsymbol{\epsilon}_j$$

其中，$\boldsymbol{\beta}$ 是常数（经常被省略，如果 $\mathbf{y}_j$ 是均值中心化后）的一个向量，$\boldsymbol{\Lambda}_y$ 是一个因子负荷矩阵，$\boldsymbol{\eta}_j$ 是含有协方差矩阵 $\boldsymbol{\Psi}$ 的 $M$ 个公共因子的一个向量，$\boldsymbol{\epsilon}_j$ 是含有对角协方差矩阵 $\boldsymbol{\Theta}$ 的独特因子的一个向量。响应的协方差矩阵变为

$$\boldsymbol{\Omega} = \boldsymbol{\Lambda}_y\boldsymbol{\Psi}\boldsymbol{\Lambda}_y' + \boldsymbol{\Theta} \tag{3.26}$$

**验证性因子分析**（confirmatory factor analysis）

如果可以得到先验信息（prior information），那么根据实体理论，以前的结论或使用的研究设计、**验证性因子分析**（CFA）就应当使用，其中，特定参数被设置为指定的数值，通常是零。例如，$\boldsymbol{\Lambda}$ 通常被设定为一个独立的群组结构（如 Jöreskog，1969；*McDonald*，1985），其中每个题器都有一个且仅有一个公共因子负荷。

例如，穆拉克（Mulaik，1988*b*）考虑了 9 个主观性等级量表的变量，它们是设计用来测量两个"维度"或因子——有关一名战士在战斗中开枪射击的构想。第一个因子，假设是**畏惧**，其指标有 4 个尺度："可怕的""神经紧张的""恐惧的"和"烦恼的"。第二个因子，**对结果乐观主义态度**，指标有 5 个尺度："有益的""有希望的""可控制的""成功的"和"可忍受的"。变量在无关因子上的负荷都被假定为零，尽管这些因子都被希望是（负）相关的。

一个独立群组双因子模型——其中每个因子都是由 3 个不重叠的题器来测量的——可写为

$$\begin{bmatrix} y_{1j} \\ y_{2j} \\ y_{3j} \\ y_{4j} \\ y_{5j} \\ y_{6j} \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \\ \beta_6 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & 1 \\ 0 & \lambda_{52} \\ 0 & \lambda_{62} \end{bmatrix} \begin{bmatrix} \eta_{1j} \\ \eta_{2j} \end{bmatrix} + \begin{bmatrix} \epsilon_{1j} \\ \epsilon_{2j} \\ \epsilon_{3j} \\ \epsilon_{4j} \\ \epsilon_{5j} \\ \epsilon_{6j} \end{bmatrix} \tag{3.27}$$

其中,我们已通过将一个因子负荷设置为1来固定每个因子尺度。该模型的一个路径图如图3.3所示,图中的圆圈表示潜变量,矩形框表示观测到的变量,连接圆圈和/或矩形框的箭头表示线性关系,指向圆圈或矩形框的短箭头表示残差变异性。连接两个变量(这里是因子)的曲线双箭头表示这两个变量是相关的。

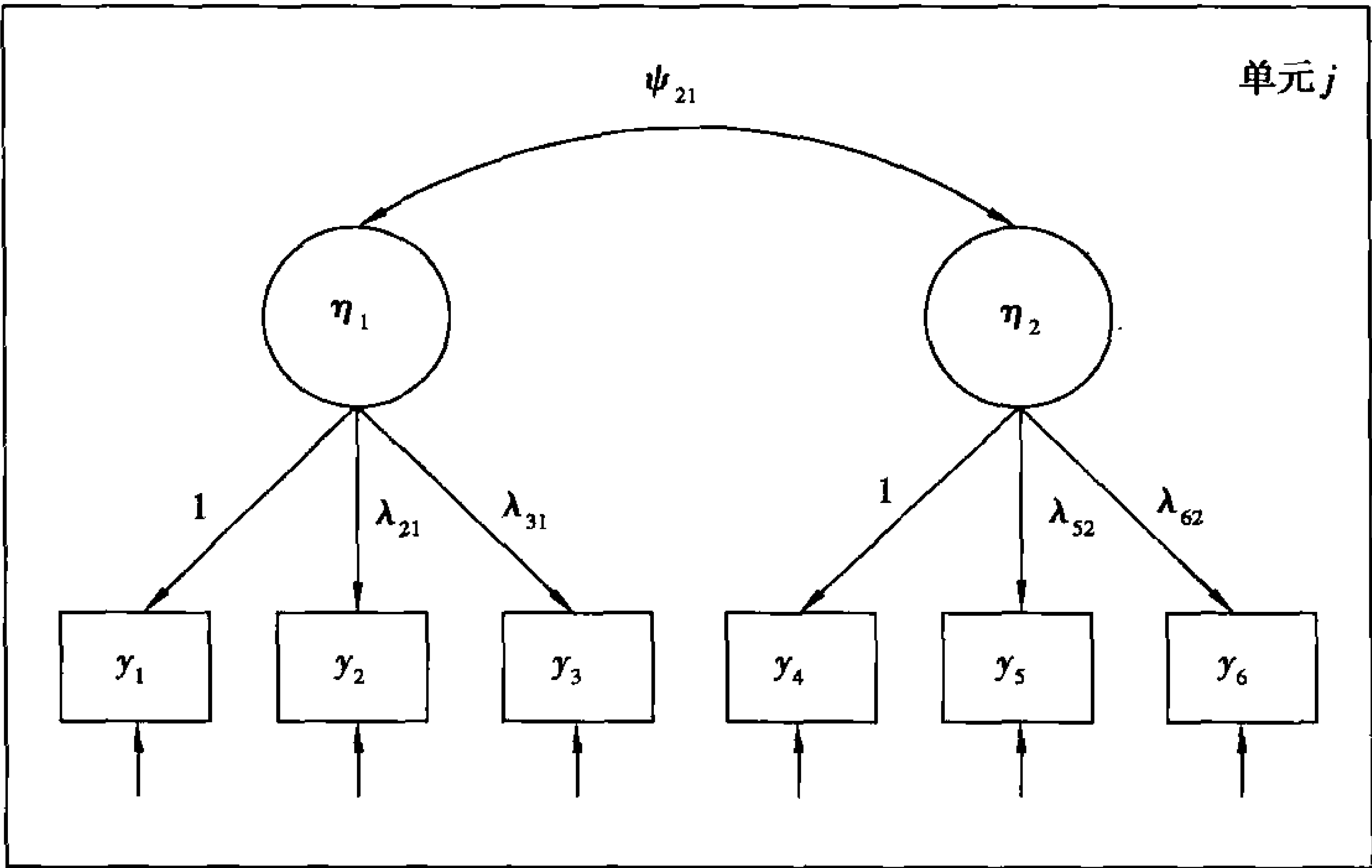



图3.3 独立群组双因子模型的路径图

通过将 $\boldsymbol{\Psi}$ 中相关的非对角元素设置为零,因子有时也被设定为不相关的。验证性因子分析因而是一个假设主义的程序,设计用来检验关于题器和因子之间关系的假设,其数量和解释已经预先决定好了。

**探索性因子分析**(Exploratory factor analysis)

另一个根本不同的方法是**探索性因子**分析(EFA)。根据穆拉克(Mulaik,1988a)的观点,探索性因子分析可以被解释为一种归纳主义的方法——设计用来**发现**一个最佳的因子集(因子的数量是在这个分析中决定的),可以解释题器之间的协变性(covariation)(参见 Holzinger and Harman,1941)。然后每个因子可以根据在因子上有高负荷的题器的子集来进行解释和"命名"。

既然我们可以把因子乘以一个任意非奇的转化矩阵 $\mathbf{R}$,即

$$\boldsymbol{\eta}_j^* = \mathbf{R}\boldsymbol{\eta}_j$$

并且可以通过把因子负荷矩阵 $\boldsymbol{\Lambda}_y$ 乘以 $\mathbf{R}^{-1}$ 来得到如式(3.26)的同样的协方差矩阵

$$\boldsymbol{\Omega} = (\boldsymbol{\Lambda}_y\mathbf{R}^{-1})\mathbf{R}\boldsymbol{\Psi}\mathbf{R}(\mathbf{R}^{-1}\boldsymbol{\Lambda}_y') + \boldsymbol{\Theta}$$

那么,如果没有对参数的若干约束,式(3.25)中的因子模型就是没有被辨识的。如果 $\mathbf{R}$ 是正交矩阵,那么,转化就是一种旋转或反射。

在验证性因子分析中,因子负荷的约束是用来固定因子旋转的,而结合对因子尺度的约束(要么通过固定因子方差,要么通过对每个因子固定一个因子负荷),则就有足够

的条件去辨识模型。在探索性因子分析中,辨识模型的一个权威但武断的方式,是设因子协方差矩阵等于单位矩阵,$\mathbf{\Psi}=\mathbf{I}$,并且对旋转进行固定,如通过要求 $\mathbf{\Lambda}_y'\mathbf{\Theta}\mathbf{\Lambda}_y$ 是对角的(如 Lawley and Maxwell,1971)。

探索性因子分析经常与**主成分分析**(principal component analysis)相混淆,因此,我们对后者进行简要的描述。主成分都是响应的线性组合 $\mathbf{a}'\mathbf{y}$,其中,$\mathbf{a}'\mathbf{a}=1$。第一个主成分的系数或"成分权重"的决定,都是用来最大化主成分的方差。每个随后的主成分系数的决定,都是用来最大化受制于这个约束——它是与前面的成分是不相关的主成分方差。响应的协方差矩阵故分解为

$$\mathrm{Cov}(\boldsymbol{y}) = \mathbf{A}\mathbf{\Psi}^{*}\mathbf{A}'$$

其中,$\mathbf{A}$ 的行是系数向量 $\boldsymbol{a}$,$\mathbf{\Psi}^{*}$ 是主成分的对角协方差矩阵。$\mathbf{A}$ 的行是 $\mathrm{Cov}(\boldsymbol{y})$ 的特征向量,$\mathbf{\Psi}^{*}$ 的对角元素是相应的特征值。与式(3.26)中的因子结构有重要差别的是,不存在独特因子协方差矩阵 $\mathbf{\Theta}$,且这些成分不能是相关的。主成分分析是一种数据简化技术,因为首要的几个主成分可能抓住了原始数据的主要特征——根据"解释方差的百分比"。与因子分析不同,没有任何一个统计模型是主成分分析的基础——它仅仅是一种对数据的转化。与主成分分析相比,因子分析的一个优势是,基于响应的不同比例(scalings),在估计之间存在一个简单的关系(如 Krane and McDonald,1978)。

通常一个探索性因子分析是通过下述相当特别的步骤来进行的。第一,因子数量是根据对相关矩阵的一个主成分分析来决定的。因子数量通常被选择等于特征值(特征值是大于 1 的)的数量,即所谓的凯瑟-古特曼(Kaiser-Guttmann)准则。然而,有时会运用到一个所谓的碎石图(scree-plot),其中,特征值对照它们的排列被划分,因子数量由曲线的"肘点"来表示(Cattell,1966)。第二,一个选好因子数量的因子模型是可以估计的。可以采用的方法很多,包括最大似然法(如 Bartholomew and Knott,1999)。然而,有些软件包实际上把主成分当做因子使用,把成分权重当做因子负荷。在这一阶段上,将意义赋予因子通常是困难的,因为大多数题器在大多数因子上都有不可忽略的负荷。第三,根据某些标准,如负荷不是"小的"就是"大的",因此,一个转化矩阵 R 可以用来产生更多可说明的负荷(如 Harman,1976;McDonald,1985)。如果不相关的因子是必要的,那么 R 必须是正交的,且这个转化是一个旋转;否则它就是指用词不当的"斜交旋转"。第四,也是最后一个步骤,是仅仅保留"显著的"负荷,将落在一个任意阈值(通常是 0.3 或0.4)之下的任何负荷解释为零。

这个最终模型的合理性从未被估计,它甚至可能不适用于那些探索性的数据。由于这个原因,加上模型建构完全是探索性的,所以我们就不会惊讶于这种模型通常是由其他样本的验证性因子分析而伪造得来的(如 Vassend and Skrondal,1995,1997,1999,2004)。我们将在 8.2.2 节回到探索性和验证性方法之间的哲学争论。

一个等价于带有 $M$ 个因子(在 5.3 节所定义的意义上)的传统探索性因子模型的验证性因子模型,可以像在探索性模型中一样,通过明智地强加 $M^2$ 个约束来进行设定。杰拉斯科格(Jöreskog,1969,1971a)的"参考解决方案"可以通过下述方法获得:

- 通过加入 $M$ 个约束 $\psi_{11}=\psi_{22}=\cdots=\psi_{MM}=1$ 来固定因子方差(在相关性上没有约束)。

- 为每个因子 $m$ 选择一个"锚"题器 $i_m$,最好它是对相应的因子带有一个大的负荷,而在其他因子上的负荷较小。对所有其他因子 $\eta_k, k\neq m$,强制 $\lambda_{i_m,k}=0$。

这是非常有用的,因为此后探索性因子分析即可通过验证性因子分析来执行,并充分利用后一方法对统计推论的便利性。我们将在 10.3 节运用这种方法来研究政治功效

的维度。

### 3.3.4 题器响应模型

一维因子模型可以通过运用两种不同的方法被扩展到二项响应和定序响应（如 Takane and de Leeuw,1987）。正如2.4节所介绍的，因子分析法通常运用一个潜响应方程。在此情况下，**潜响应** $y_{ij}^*$ 则简单地取代了传统因子模型中的观测响应 $y_{ij}$。另一方面，在题器响应理论（IRT）中，广义线性模型方程是经常被使用的。这里，一个已知潜特质（或因子）的特定响应，它的条件概率（即所谓的"题器特质曲线"）是通过一个链接函数（通常是一个对数概率或正态概率链接）来设定的。

在这个方程式中，单个因子模型是大家所熟悉的双参数题器响应模型，因为与每个题器相联系的有两个参数，即一个截距和一个因子负荷。这些模型的经典应用是在能力测试中，其中，题器 $i$ 代表一个测试中的题目或问题，回答被记录为对(1)或错(0)。在这个设置中，$\eta_j$ 代表被试人 $j$ 的能力，模型通常可以被参数化为

$$\nu_{ij} = a_i(\eta_j - b_i)$$

其中，$b_i$ 可以被理解为试题难度，当能力等于难度时，给出一个50%的机会作出正确答案，而 $a_i$ 是一个题器区分度参数（或因子负荷），决定了题器在拥有不同能力被试对象之间的区别有多好。如图3.4所示的例子展示了一个双参数对数概率（2-PL）题器响应模型的题器特质曲线（Birnbaum,1968）。实线和虚线代表的是具有同样难度 $b$ 但区分度参数（斜率）$a$ 不相同的题器，而实线和圆点曲线代表的是具有同样的区分度参数 $a$ 但难度 $b$ 不相同（水平移动）的题器。我们将在9.4节中估计一个关于测试数学推理的题器的双参数题器响应模型。

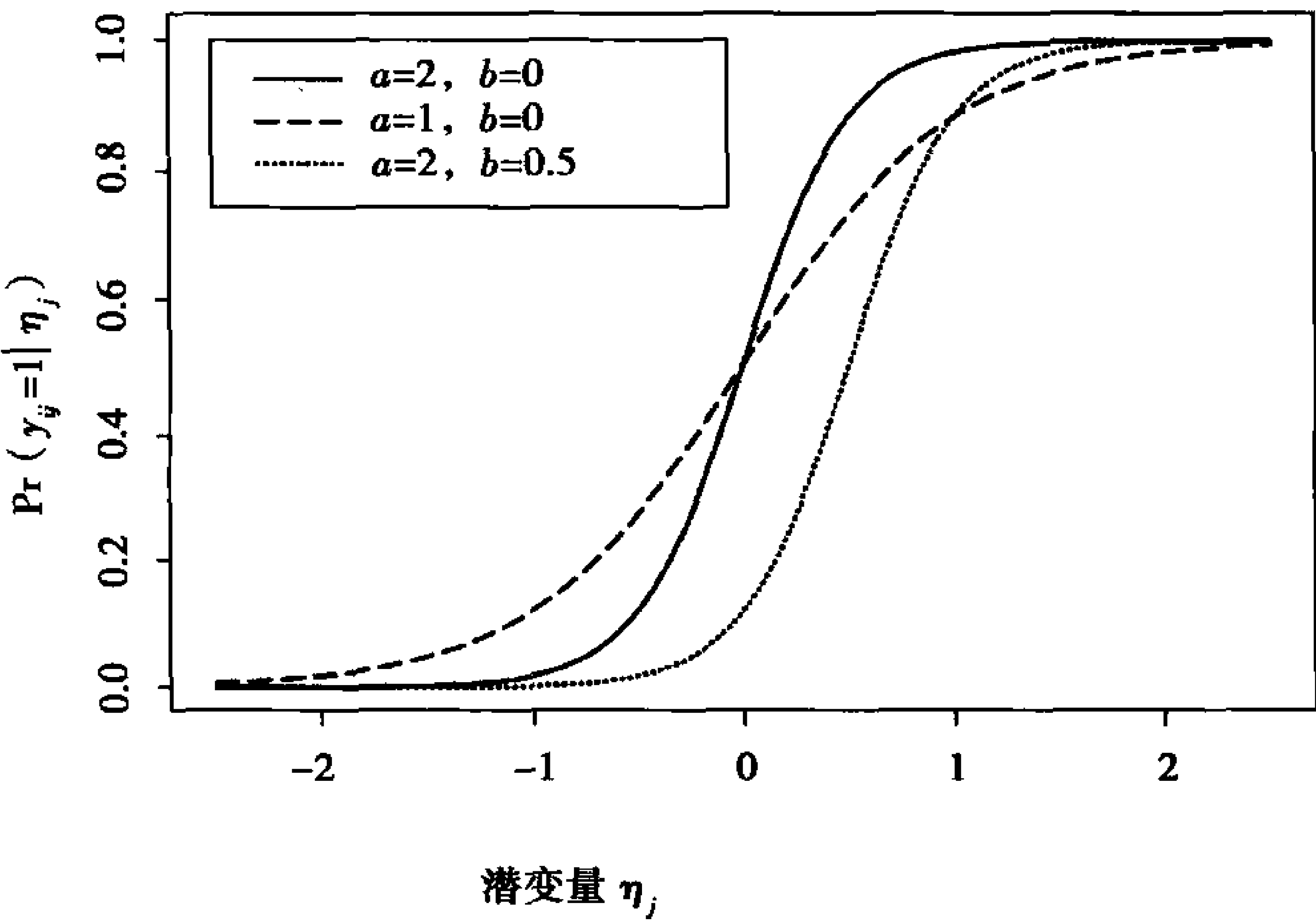



**图3.4 具有不同 $a$ 和 $b$ 的双参数对数概率模型的题器特质曲线**

在双参数模型中，当能力趋于负无穷大时，正确回答的概率趋于零。然而，如果使用了多项选择题型，那么这就是不切实际的，因为猜测会产生一个正确回答的非零概率，即使是对那些能力低下的人。因此，有时会将一个额外参数引入双参数模型中，从而形成三参数对数概率题器响应模型（Birnbaum,1968），参见9.4节的一个例子。遗憾的是，为了获得对这一模型的可靠估计，大样本看来是必要的（如 Wainer and Thissen,1982）。我们认为这一问题在麦当劳（McDonald,1967）提出的四参数模型中会更加恶化。该模型引进一个额外参数来捕获那些甚至是极有能力的应试者也会犯的错误。

在双参数题器响应模型中，因子负荷 $a_i$ 通常被强制为相等的，并且不失一般性地设

置为1,得到

$$\nu_{ij} = \eta_j - b_i$$

这是一个单参数模型。在对数概率的情形中,这个单参数模型经常被简写为1-PL。注意,一个单参数题器响应模型,对于没有协变量的二项题器来说,正好是一个随机截距模型。从图3.5中可以明显看出,单参数模型有个特性——有时称为"双重单调性(double-monotonicity)":对于每个能力,其表现是随着难度降低而减少的;对于每个难度,其表现随着能力的提高而增加,即题器和测试对象是严格有序的。

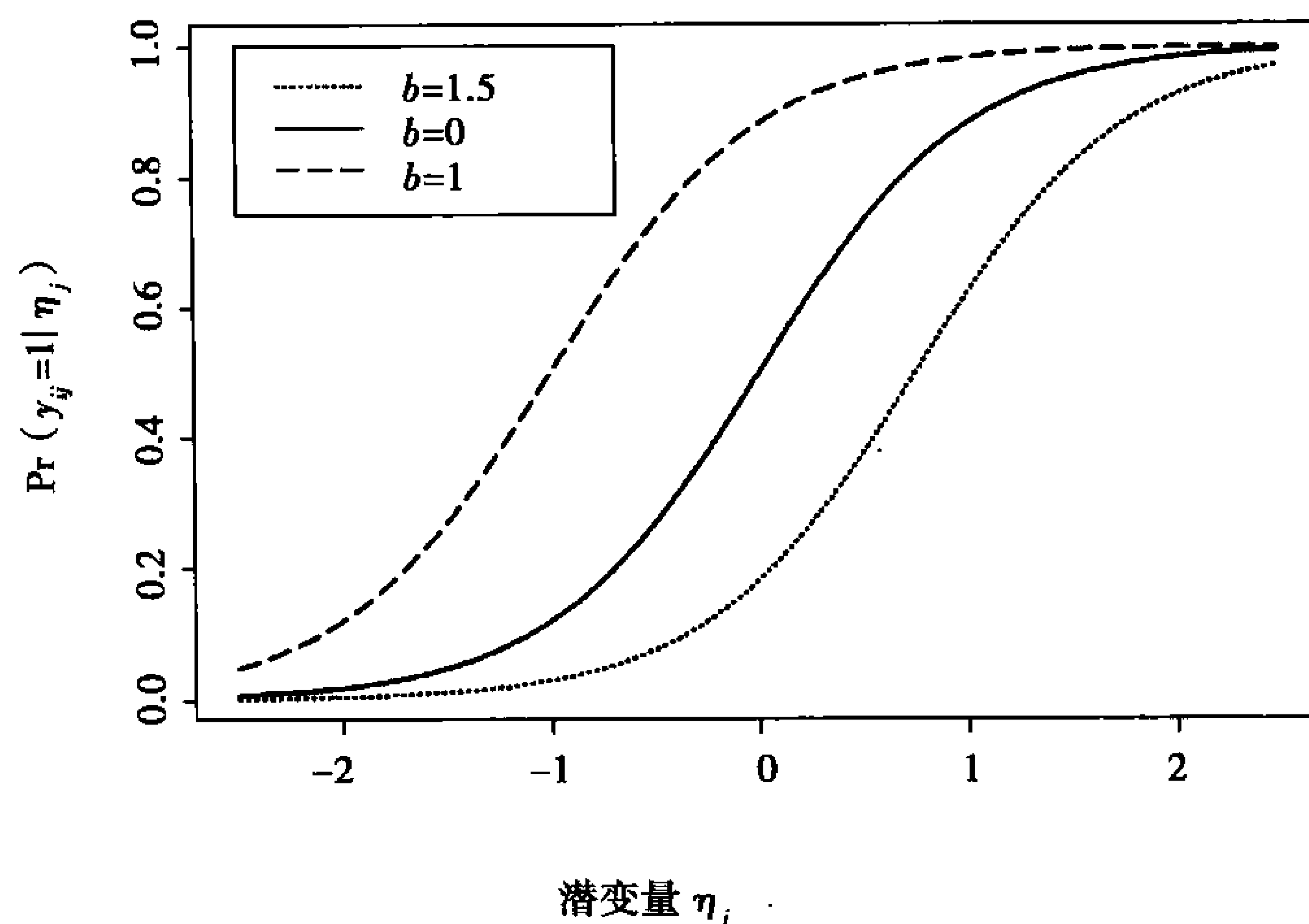



**图3.5 具有不同 $b$ 的单参数对数概率模型的题器特质曲线**

如果单参数模型中的 $\eta_j$ 被看做是固定参数,并运用一个对数概率链接,那么就得到著名的拉什模型(Rasch,1960)。该模型具有许多引人注目的理论品质(如Fischer,1995)。例如,拉什模型等价于这一必要条件,即对于给定题器参数 $b_i$ 的 $\eta_j$ 来说,响应的未加权总得分是一个充分统计量。这意味着条件最大似然估计能够被应用到题器参数中(参见6.10.2节)。此外,拉什模型是一个具有科学论述之普遍性的特定概念,由拉什(Rasch,1967)命名为"特定客观性(specific objectivity)"。一般来说,特定客观性是指两个测试对象的能力的比较,仅仅是依赖于这两个对象的能力(而不是其他人的能力),并且这个比较应当服从同一结果,无论这个比较是基于哪个题器的。

作为对拉什模型的一个发展,费希尔(Fischer,1977,1995)提出了**线性对数概率检验模型**,其中,拉什模型中的题器参数是根据特定题器的协变量 $\mathbf{x}_i$ 来构建的,即

$$b_i = \mathbf{x}_i'\boldsymbol{\beta}$$

定类响应模型(nominal response model)(如Rasch,1961;Bock,1972;Samejima,1972;Andersen,1973)处理了多项响应,比如多项选择题的答案 $a_s = 1, \cdots, S$。它是一个多元对数概率模型,即

$$\Pr(y_{ij} = a_s) = \frac{\exp(\nu_{ij}^s)}{\sum_{t=1}^{S_i} \exp(\nu_{ij}^t)}$$

其中

$$\nu_{ij}^s = \beta_i^s + \lambda_i^s \eta_j$$

模型的辨识是可实现的,例如,通过设 $\beta_i^1 = 0, \lambda_i^1 = 0$ 以及 $\mathrm{Var}(\eta_j) = 1$。

在题器响应理论中,定序响应的著名模型,包括**部分评分模型**(partial credit model)(Masters,1982)和**等级尺度模型**(rating scale model)(Andrich,1978),可以通过在定类响

应模型中强加约束而获得(如 Thissen and Steinberg,1986)。在部分评分模型中,线性估计量被给定为

$$\nu_{ij}^{s} = \beta_{i}^{s} + s\eta_{j}$$

其中,对于定类响应模型的未知因子负荷来说,等距的类别得分 $s$ 是可替代的。差分 $\Delta_{i}^{s} = -(\beta_{i}^{s} - \beta_{i}^{s-1})$ 有时被称为与类别 $s$ 相关联的"题器步进难度"(item step difficulty)。

等级尺度模型是部分评分模型的一种特殊情况,其中

$$\nu_{ij}^{s} = (\delta_{i} + \tau^{s}) + s\eta_{j}$$

截距被划分成特定题器成分 $\delta_{i}$ 和特定类别成分 $\tau^{s}$。

## 3.4 潜类模型

在潜类模型中,单元是被假设属于 $C$ 个离散类别($c = 1,\cdots,C$)中的一个,其中类别隶属是未知的。因此,类别可以被视为一个分类潜变量的种类。一个单元 $j$ 属于类别 $c$ 的(先验)概率 $\pi_{jc}$,是一项模型参数。

潜类模型传统上应用于这种情况,即二项响应或多项响应 $i$ 对于每个单元 $j$ 都是可观测到的。如果单元 $j$ 在类别 $c$ 中,那么题器 $i$ 取值为 $a_{s}(s = 1,\cdots,S_{i})$ 的条件响应概率,可被构建为一个多元对数概率模型,即

$$\Pr(y_{ij} = a_{s} \mid c) = \frac{\exp(\nu_{ijc}^{s})}{\sum_{t=1}^{S}\exp(\nu_{ijc}^{t})}$$

题器的响应都被假设为条件独立于一个给定潜类别中的给定隶属关系。这与题器响应模型和因子模型是相似的,其中响应都是条件独立于给定的一个连续性潜特质。

无条件响应概率成为有限混合(finite mixtures)形式

$$\Pr(y_{ij} = a_{s}) = \sum_{c=1}^{C}\pi_{jc}\Pr(y_{ij} = a_{s} \mid c)$$

而一种响应模式 $\mathbf{y}_{j} = (y_{1j},\cdots,y_{Ij})$ 的概率为

$$\Pr(\mathbf{y}_{j}) = \sum_{c=1}^{C}\pi_{jc}\prod_{i=1}^{I}\Pr(y_{ij} \mid c)$$

在传统探索性潜类模型中(如 McCutcheon,1987),题器 $i$ 和种类 $s$ 的线性估计量是一个自由参数,即

$$\nu_{ijc}^{s} = \beta_{ic}^{s}$$

带有一个约束 $\beta_{ic}^{1} = 0$。我们将在 9.3 节估计一个探索性潜类模型——关于心肌梗塞的 4 个二项诊断测试。两个类别分别代表患有心肌梗塞和未患心肌梗塞的病人,线性估计量中的参数与测试本身的灵敏度和特异性有关。

就定序响应而言,对某些得分 $b_{s}$,$\beta_{ic}^{s}$ 经常被构造为 $\beta_{ic}^{s} = b_{s}\beta_{c}$,当 $b_{s} = s$ 时得到邻类对数概率模型。其他参数化形式也是可能的,参见 2.3.4 节。验证性潜类模型对参数强加一些约束,通常是设定某些条件响应概率等于交叉潜类别。潜类模型可以被公式化为处理列联表的对数线性模型,其中类别变量之一(即潜类变量)是未观测到的(如 Goodman,1974;Haberman,1979)。

有限混合模型具有与潜类模型相同的结构。对于连续响应和计数,当每个单元仅有一个响应时,这些模型经常会被使用,以获得一个概率分布的灵活模型。在多元连续响应的情况下,条件响应分布经常被设定为多元正态分布,从而释放常见的条件独立假设。

这种基于模型的群组分析在班菲尔德和拉夫特里(Banfield and Raftery,1993)以及本斯梅尔等人(Bensmail et al. ,1997)的著作中有所讨论。

## 3.5 带有潜变量的结构方程模型

测量模型和因子模型在它们自身领域是非常重要的,并且也是结构方程模型中的基础部分,其中潜变量之间的关系是模型化的。这些关系通常具有主要的科学价值,而观测题器与潜变量之间的关系具有次要价值。直接对潜变量之间的关系进行建模的一个重要优势是,测量误差的不利影响,如回归稀释(regression dilution)(如 Rosner et al. , 1990),可能会潜在地被修正(参考 Fuller,1987 and Carroll et al. ,1995a)。

现来考虑简单的“变量含误差”问题,其中一个单一协变量 $\xi_j$ 的测量是带有误差的,依据的是一个传统测量模型

$$x_j = \xi_j + \delta_j \tag{3.28}$$

信度 $\rho<1$。我们希望研究 $y_j$ 在真实协变量(理想得分)$\xi_j$ 上的回归

$$y_j = \gamma_0 + \gamma_1\xi_j + \zeta_j \tag{3.29}$$

并对回归参数 $\gamma_1$ 特别感兴趣。然而,使用传统回归,我们必须依赖于这个回归是在观测协变量上的,但它通常由易于出错的协变量 $x_j$ 所代替,即

$$y_j = \gamma_0^* + \gamma_1^* x_j + \zeta_j^*$$

在这个简单情况下,可以看到忽略测量误差的结果是,估计的回归参数被削弱了与真实回归参数的关系

$$\mathbf{E}(\hat{\gamma}_1^*) = \gamma_1\rho$$

当存在几个带有测量误差的协变量时,这些结果就不那么清晰了。

重要的是,$\gamma_1$ 可以一直通过对式(3.28)和式(3.29)的联合建模来进行估计,这是结构方程模型的一个简单例子。同时还应当注意,在伯克森测量模型中或者当仅仅是响应变量带有测量误差时,衰减(attenuation)问题不会出现。

现在我们介绍带有潜变量的传统结构方程模型,也经常被指称为协方差结构分析。正如后一术语所暗示的,其关注的焦点是协方差结构,而均值结构通常是被消去的——通过从每个变量中减去均值。由于已经定义了公共因子模型,因此,一个设定潜变量之间关系的结构模型就可以被建构。在这个结构模型中,可能既有潜在因变量,也有潜在解释变量。例如,考虑一个针对两个潜在因变量 $\eta_{1j}$和 $\eta_{2j}$以及两个潜在解释变量 $\xi_{1j}$和 $\xi_{2j}$ 的结构方程模型。因变量的测量模型被设定为一个独立群组模型,其中,每个潜在因变量的测量都是通过 3 个不重叠的题器(如式(3.27))来进行的,用向量符号可写为

$$\mathbf{y}_j = \mathbf{\Lambda}_y\boldsymbol{\eta}_j + \boldsymbol{\epsilon}_j$$

类似地,解释变量的测量模型可写为

$$\mathbf{x}_j = \mathbf{\Lambda}_x\boldsymbol{\xi}_j + \boldsymbol{\delta}_j$$

注意我们已经省略了常数,假定 $\mathbf{y}_j$ 和 $\mathbf{x}_j$ 都以均值为中心的。

现在我们来设定一个结构模型,让两个潜在因变量在两个潜在解释变量上进行回归。另外,一个潜在因变量在另一个上回归,则

$$\begin{bmatrix}\eta_{1j}\\ \eta_{2j}\end{bmatrix} = \begin{bmatrix}0 & 0\\ \beta_{21} & 0\end{bmatrix}\begin{bmatrix}\eta_{1j}\\ \eta_{2j}\end{bmatrix} + \begin{bmatrix}\gamma_{11} & \gamma_{12}\\ \gamma_{21} & \gamma_{22}\end{bmatrix}\begin{bmatrix}\xi_{1j}\\ \xi_{2j}\end{bmatrix} + \begin{bmatrix}\zeta_{1j}\\ \zeta_{2j}\end{bmatrix}$$

这是**递归模型**(recursive model)的一个例子,其中,潜在因变量之间的倒数关系或反馈关

系是不存在的。用更一般的形式,这个模型可变为

$$\boldsymbol{\eta}_j = \mathbf{B}\boldsymbol{\eta}_j + \boldsymbol{\Gamma}\boldsymbol{\xi}_j + \boldsymbol{\zeta}_j \tag{3.30}$$

结构方程模型经常表示为路径图的形式,图3.6中给出一个例子描述了这一模型,假定$\xi_{1j}$和$\xi_{2j}$是相关的,而$\zeta_{1j}$和$\zeta_{2j}$是不相关的。

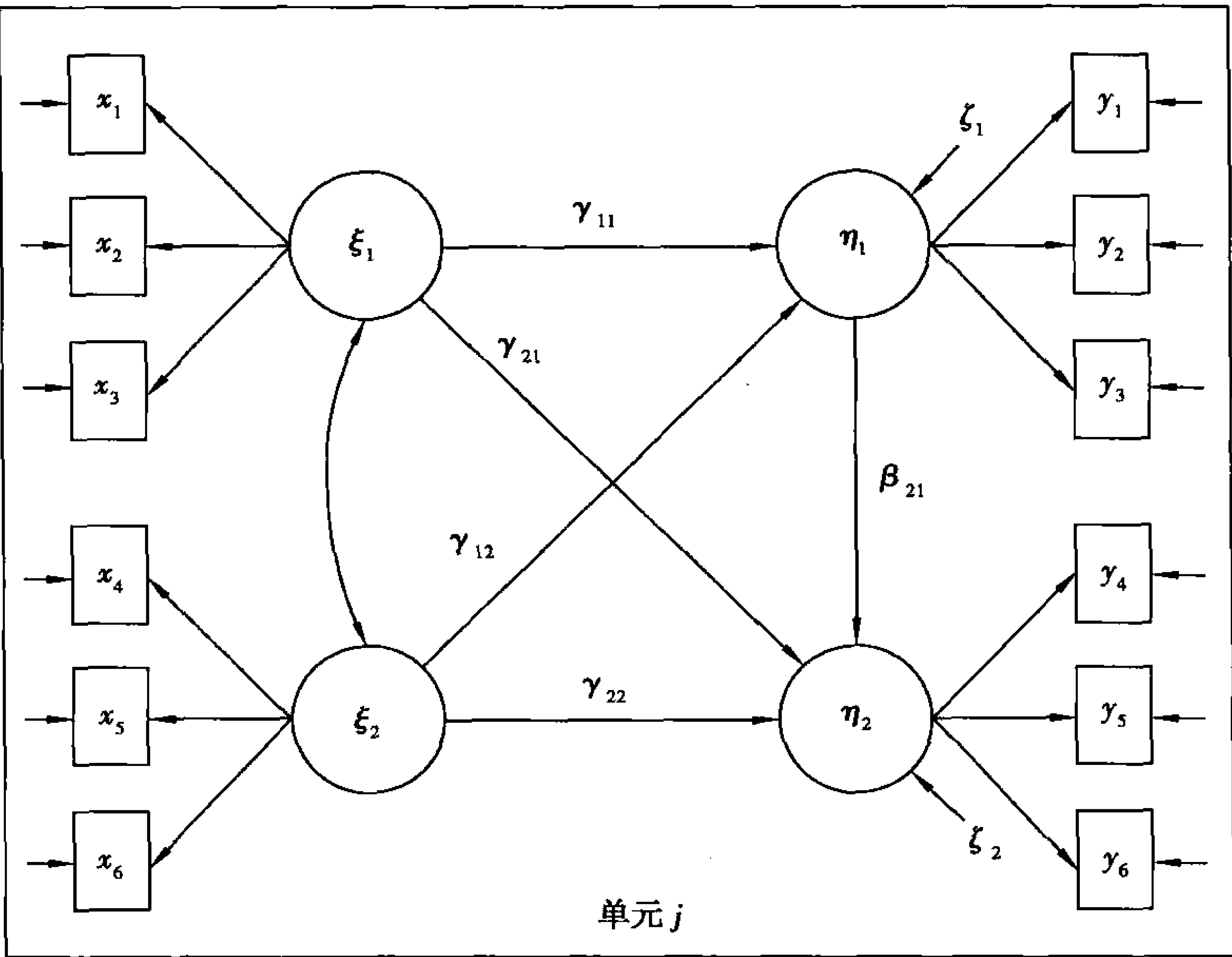



**图3.6 结构方程模型的路径图**

上述LISREL模型(Jöreskog,1973)是带有潜变量的结构方程模型的支配性参数化形式,但是几种另外的参数化也已经提出(参见Bollen,1989:395)。有时,所有测量都被叠入一个长向量$\mathbf{y}_j$中,而潜变量相应地叠入一个单一向量$\boldsymbol{\eta}_j$中,误差项叠入$\boldsymbol{\epsilon}_j$中。然后这个组合的测量模型可写为

$$\mathbf{y}_j = \boldsymbol{\Lambda}\boldsymbol{\eta}_j + \boldsymbol{\epsilon}_j \tag{3.31}$$

其中

$$\boldsymbol{\Lambda} \equiv \begin{bmatrix} \boldsymbol{\Lambda}_y & 0 \\ 0 & \boldsymbol{\Lambda}_x \end{bmatrix}$$

结构模型变为

$$\begin{bmatrix} \eta_{1j} \\ \eta_{2j} \\ \xi_{1j} \\ \xi_{2j} \end{bmatrix} = \begin{bmatrix} 0 & 0 & \gamma_{11} & \gamma_{12} \\ \beta_{21} & 0 & \gamma_{21} & \gamma_{22} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \eta_{1j} \\ \eta_{2j} \\ \xi_{1j} \\ \xi_{2j} \end{bmatrix} + \begin{bmatrix} \zeta_{1j} \\ \zeta_{2j} \\ \zeta_{3j} \\ \zeta_{4j} \end{bmatrix}$$

也可简洁地写为

$$\boldsymbol{\eta}_j = \mathbf{B}\boldsymbol{\eta}_j + \boldsymbol{\zeta}_j \tag{3.32}$$

如果存在一个观测协变量,那么,通过引入一个模拟的"潜在"解释变量,它就可以被包括到模型当中,对于该协变量,因子负荷等于1;对于所有其他观测变量,因子负荷为0;并且将独特因子方差设为0。这个方法的一个缺点是,解释变量事实上被当做是响应变量,以至于要经常借助于多元正态假设。这对于很多连续的协变量来说,明显是不合理的,甚至对于像性别这样的二项协变量也是如此。有些架构(如Muthén,1984)为观

测协变量上的潜变量回归纳入一个额外项 $\boldsymbol{\Gamma}\mathbf{x}_{1j}$：

$$\boldsymbol{\eta}_j = \boldsymbol{\alpha} + \mathbf{B}\boldsymbol{\eta}_j + \boldsymbol{\Gamma}\mathbf{x}_{1j} + \boldsymbol{\zeta}_j \tag{3.33}$$

其中，$\boldsymbol{\alpha}$ 是一个截距向量。穆森设定该模型条件依赖于协变量，从而使这些协变量不再需要分布假设。在这个测量模型中，穆森和穆森（Muthén and Muthén，1998）纳入了一个附加项 $\mathbf{K}\mathbf{x}_{2j}$来表示观测响应在观测协变量上的回归，即

$$\mathbf{y}_j = \boldsymbol{\nu} + \boldsymbol{\Lambda}\boldsymbol{\eta}_j + \mathbf{K}\mathbf{x}_{2j} + \boldsymbol{\epsilon}_j \tag{3.34}$$

其中，$\boldsymbol{\nu}$ 是截距的一个向量（通常 $\mathbf{x}_{1j} = \mathbf{x}_{2j}$）。

一个普及的带有观测协变量的结构方程模型是多指标多原因（Multiple-Indicator Multiple-Cause）（MIMIC）模型，它是一个单因子模型，其中，因子是由多项指标来测量的，并且在几个观测协变量或"原因"上回归（如 Zellner，1970；Hauser and Goldberger，1971；Goldberger，1972）。此时结构模型简单是

$$\eta_j = \alpha + \boldsymbol{\gamma}'\mathbf{x}_{1j} + \zeta_j$$

一个带有 3 个指标和 3 个协变量的 MIMIC 模型的路径图如图 3.7 所示。

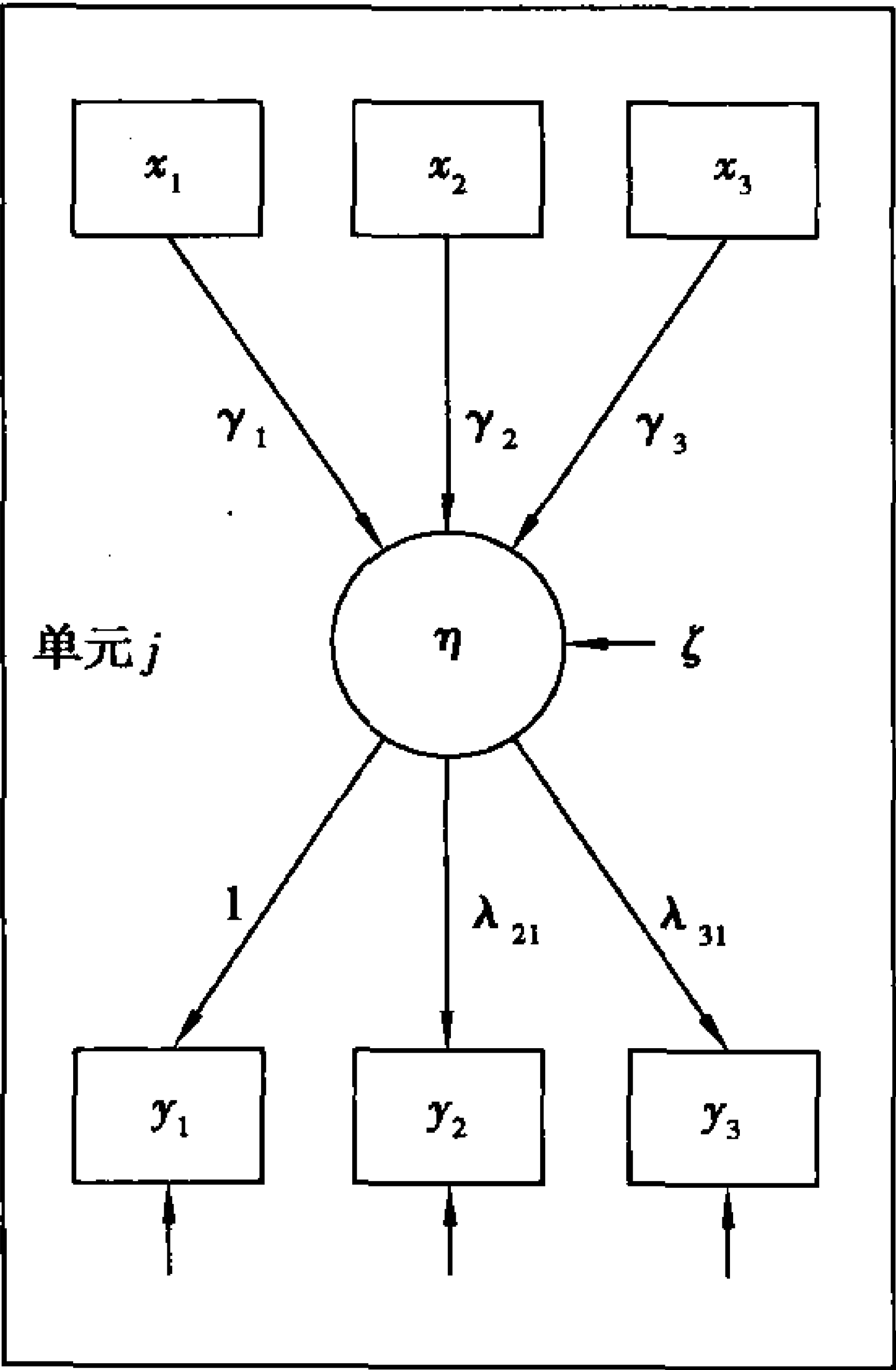



**图 3.7 MIMIC 模型的路径图**

罗宾斯和韦斯特（Robins and West，1977）考虑了用一个 MIMIC 模型来处理房屋价值估计中的测量误差。在模型的测量部分使用了 3 种房屋价值测量方法："由私人公司来评估价值""由所有者来估计价值""由征税的县来估价"。在这个结构部分，房屋价值在 12 种财产特征上进行回归，包括"建筑等级""车库类型"和"完成面积"。关于带有几个因子的 MIMIC 模型，我们可参考 Robinson（1974）。

回到方程（3.33）中的一般模型，这个结构模型可以被解答，因为潜变量给出

$$\boldsymbol{\eta}_j = (\mathbf{I} - \mathbf{B})^{-1}[\alpha + \boldsymbol{\Gamma}\mathbf{x}_{1j} + \boldsymbol{\zeta}_j] \tag{3.35}$$

将 $\boldsymbol{\eta}_j$ 代入方程（3.34），得到简化形式

$$\mathbf{y}_j = \boldsymbol{\nu} + \boldsymbol{\Lambda}(\mathbf{I} - \mathbf{B})^{-1}[\boldsymbol{\alpha} + \boldsymbol{\Gamma}\mathbf{x}_{1j} + \zeta_j] + \mathbf{K}\mathbf{x}_{2j} + \boldsymbol{\epsilon}_j$$

已知 $\mathbf{x}_{1j}$和 $\mathbf{x}_{2j}$的条件期望结构变为

$$\mathbf{E}(\mathbf{y}_j \mid \mathbf{x}_{1j}, \mathbf{x}_{2j}) = \boldsymbol{\nu} + \boldsymbol{\Lambda}(\mathbf{I} - \mathbf{B})^{-1}[\boldsymbol{\alpha} + \boldsymbol{\Gamma}\mathbf{x}_{1j}] + \mathbf{K}\mathbf{x}_{2j}$$

且条件协方差结构为

$$\boldsymbol{\Omega} \equiv \mathrm{Cov}(\mathbf{y}_j \mid \mathbf{x}_{1j}, \mathbf{x}_{2j}) = \boldsymbol{\Lambda}(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Psi}(\mathbf{I} - \mathbf{B})^{-1\prime}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}$$

其中,$\Psi$ 是 $\zeta_j$ 的协方差矩阵,$\Theta$ 是 $\epsilon_j$ 的协方差矩阵。参见第 5 章关于带有潜变量的传统结构方程模型的深入案例。

结构方程模型也经常被给予因果解释。如,戈德伯格(Goldberger,1972)在这种情境中谈论"因果联系(causal links)",并且这种方法论以前一直被称为"因果建模(causal modeling)"(如 Bentler,1978,1980;James et al.,1982)。附于联立方程(simultaneous equations)中的因果性用法,毫无疑问地既是这些模型在社会科学家之间具有吸引力的一个主要原因,也是导致统计学家持怀疑态度的主因之一。因果解释在观测设计的情境中当然应该极其谨慎地进行操作,强调这一点还有 Guttmann(1977)、Cliff(1983)、Freedman(1985,1986,1992)、Holland(1988)、Sobel(1995)。

结构方程模型的设定和相应路径图的制作,对于推理因果过程如何进行,仍然是不可缺少的。例如,在流行病学中,一个简单的路径图通常可以揭示哪些变量最好被视为"混淆因子(confounder)"(且需要"控制的"),以及哪些变量应当被视为中间变量(且不需要"控制的")。注意此处讨论的结构方程模型是与图模型(graphical model)和针对潜在后果的模型是密切相关的,这一点很重要(参见 Greenland and Brumback,2002)。珀尔(Pearl,2000)提供了一个明晰的关于现代"因果建模"的论述。

## 3.6 纵贯性模型

纵贯性数据,在医学中通常称为重复测量,在社会科学中称为面板数据,在经济学中称为截面时间序列数据(cross-sectional time-series data),出现于当单元在多个时刻上作出响应时。纵贯性数据的两个重要特征是响应在单元内的聚集和响应按时间顺序排列。一个典型的问题是去研究全部响应层次的估计量和响应随时间变化的估计量。正如 1.4 节所论证的,纵贯性设计允许截面效应与纵贯性效应的分离。

除了调节均值结构,纵贯性模型还必须考虑到同一个单元中响应之间的相关性。这一相关性的原因包括:单元之间未观测到的异质性——引起单元内的相关性(如在双层次模型中),以及未观测到的时间变动效应——导致在发生时间上靠得更近的响应之间有更大的相关性。这两种类型的未观测到的异质性,都能够明确地被模型化,以便尝试去解释条件协方差结构(已知协变量)。

在随后的小节中,我们将对两种类型的纵贯性数据进行辨析:**对称**时点(balanced occasions)的数据或**不对称**时点(unbalanced occasions)的数据。如果所有单元都是在同一时点集 $t_i(i=1,\cdots,I)$ 上被测量,那么时点就是对称的;如果不同单元的测量是在不同的时点集 $t_{ij}(i=1,\cdots,n_j)$ 上进行的,那么时点就是不对称的。在任一情况中,都可能会出现缺失数据。如果不同单元是在不同的时点集上进行测量,但是在每一个时间点上,都能得到相当多的单元的测量,那么时点也可以被视为带有缺失数据的对称时点。

不论对于对称时点还是不对称时点,纵贯性数据都可以被认为是双层次数据,时点 $i$ 在第一层次,单元 $j$ 在第二层次。就对称时点而言,数据还可以被视为单层次多元数据,其中,不同时点上的响应都被视为不同变量。在这一情况下,均值结构和协方差结构的模型可以纳入特定时点的参数,如特定时点的残方差。然而,就不对称时点而言,均值和协方差结构通常被模型化为一个与时点相联系的时间函数,或者是一个时间变动协变量的函数。

### 3.6.1 特定单元效应模型

为每个单元在多个时点上收集信息的主要原因是，它能达到对个体单元内的变化的调查研究；特定单元（在时间上是恒定的）效应能够被控制，并且我们可以研究单元间变异性的特征、变化的程度或是随着时间变化的增长情况。也许，纵贯性建模最通常的方法就是建构个体增长轨迹模型，运用一个公共（单元间交叉）固定效应来概括轨迹的平均特征，或者用特定单元的固定或随机效应去表示单元间的变异性。于是，随机效应就导致并进而解释了条件协方差结构。

**固定效应模型**

考虑单元$j$在时点$i$上的响应$y_{ij}$。纵贯性数据的一个简单线性固定效应模型是如下形式

$$y_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \alpha_j + \epsilon_{ij} \tag{3.36}$$

其中，$\mathbf{x}_{ij}$是带有回归参数$\boldsymbol{\beta}$的时间变动协变量（有时包括一个时间变量$t_{ij}$），$\alpha_j$是特定单元截距或"固定效应"，$\epsilon_{ij}$是同一性的和独立的正态分布的残差，其中$E(\epsilon_{ij}|\mathbf{x}_{ij})=0$。固定效应$\alpha_j$代表特定单元的效应，如果被忽略，可能会导致混淆并引起残差之间的相关性，从而产生偏差。

作为在模型中纳入$\alpha_j$的一个结果，效应$\boldsymbol{\beta}$能够被解释为单元间效应。这一点可以通过考察模型(3.36)中的群组均值来发现，则

$$\bar{y}_j = \bar{\mathbf{x}}'_j\boldsymbol{\beta} + \alpha_j + \bar{\epsilon}_j \tag{3.37}$$

其中，响应都是单元（在各时点之上）的均值。由于每个响应都对应一个单独的参数$\alpha_j$（得到一个饱和模型），因此，响应并不对$\boldsymbol{\beta}$提供任何信息。用式(3.36)减去式(3.37)，得到了单元内回归模型为

$$y_{ij} - \bar{y}_j = (\mathbf{x}_{ij} - \bar{\mathbf{x}}_j)'\boldsymbol{\beta} + \epsilon_{ij} - \bar{\epsilon}_j \tag{3.38}$$

消掉了$\alpha_j$。

对$\boldsymbol{\beta}$和$\alpha_j$的估计可以从对固定效应模型(3.36)的普通最小二乘(OLS)估计中获得，它同时产生了两个估计$\hat{\boldsymbol{\beta}}_{FE}$和$\hat{\alpha}_j$。或者等价地，对$\boldsymbol{\beta}$的OLS估计可以建立在单元内模型(3.38)的基础上，产生

$$\hat{\boldsymbol{\beta}}_W = W_{xx}^{-1} W_{xy}$$

其中，$W_{xx} \equiv \sum_{i,j}(\mathbf{x}_{ij} - \bar{\mathbf{x}}_j)(\mathbf{x}_{ij} - \bar{\mathbf{x}}_j)'$，$W_{xy} \equiv \sum_{i,j}(\mathbf{x}_{ij} - \bar{\mathbf{x}}_j)(y_{ij} - \bar{y}_j)$。于是，特定单元截距则可被估计为

$$\hat{\alpha}_j = \bar{y}_j - \mathbf{x}'_{ij}\hat{\boldsymbol{\beta}}_W$$

如果假设$\epsilon_{ij}$是正态分布的，那么$\hat{\boldsymbol{\beta}}_W$也可以从已知$\sum_{i=1}^{n} y_{ij}$的条件最大似然估计获得。实际上，当时点的数量$n$是固定的并且属于一个适中的量值，$\hat{\boldsymbol{\beta}}_W$是一个最好的无偏估计量，并且是前后一致的，相反，当$N\to\infty$时，$\hat{\alpha}_j$则是不一致的。

在单变量重复测量ANOVA中，或者在对称的裂区设计(split plot design)的ANOVA中（如Hand and Crowder, 1996），单元内效应可以通过上述模型进行估计，而单元间效应可以通过为均值响应$\bar{y}_j$设定一个普通线性模型进行估计。第一个分析相当于对单元内平方和的一种分块，而第二个分析则是运用单元间的平方和。

展示 3.3 纵贯性数据的公共相关结构

A. 随机截距结构：

$$\boldsymbol{\Omega} = \psi\mathbf{1}\mathbf{1}' + \theta\mathbf{I} = \begin{bmatrix} \psi+\theta & & & \\ \psi & \psi+\theta & & \\ \vdots & \vdots & \ddots & \\ \psi & \psi & \cdots & \psi+\theta \end{bmatrix}$$

B. 随机系数结构：

$$\boldsymbol{\Omega}_j = \mathbf{Z}_j\boldsymbol{\Psi}\mathbf{Z}_j' + \theta\mathbf{I}_{N_j}$$

C. 自回归的(Autoregressive)残差结构 AR(1)：

$$\boldsymbol{\Omega} = \frac{\sigma_\delta^2}{1-\alpha^2}\begin{bmatrix} 1 & & & \\ \alpha & 1 & & \\ \vdots & \vdots & \ddots & \\ \alpha^{I-1} & \alpha^{I-2} & \cdots & 1 \end{bmatrix}$$

D. 移动平均残差结构 MA(1)：

$$\boldsymbol{\Omega} = \sigma_\delta^2\begin{bmatrix} 1+a^2 & & & & \\ a & 1+a^2 & & & \\ 0 & a & 1+a^2 & & \\ \vdots & \vdots & \vdots & \ddots & \\ 0 & 0 & 0 & \cdots & 1+a^2 \end{bmatrix}$$

E. 自回归的响应结构 AR(1)：

$$\boldsymbol{\Omega} = \frac{\sigma_\epsilon^2}{1-\gamma^2}\begin{bmatrix} 1 & & & \\ \gamma & 1 & & \\ \vdots & \vdots & \ddots & \\ \gamma^{I-1} & \gamma^{I-2} & \cdots & 1 \end{bmatrix}$$

F. 因子结构：

i. 一个因子：

$$\boldsymbol{\Omega} = \boldsymbol{\Lambda}\psi\boldsymbol{\Lambda}' + \boldsymbol{\Theta} = \begin{bmatrix} \lambda_1^2\psi+\theta_{11} & & & \\ \lambda_2\psi\lambda_1 & \lambda_2^2\psi+\theta_{22} & & \\ \vdots & \vdots & \ddots & \\ \lambda_I\psi\lambda_1 & \lambda_I\psi\lambda_2 & \cdots & \lambda_I^2\psi+\theta_{II} \end{bmatrix}$$

ii：多维因子：

$$\boldsymbol{\Omega} = \boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}$$

固定效应模型的一个问题是，时间常数协变量(如性别或治疗组)的回归参数(其中，$\mathbf{x}_{ij}=\bar{\mathbf{x}}_j$)是未被辨识的，参见式(3.38)。固定效应模型的条件最大似然估计在6.10.2节也有所讨论。

**随机截距模型**

代替将特定单元效应看做是固定的，我们可以假设这些效应是一个随机变量 $\zeta_j$ 的实现结果，即

$$y_{ij} = \mathbf{x}_{ij}'\boldsymbol{\beta} + \zeta_j + \epsilon_{ij}$$

其中，$\zeta_j$ 和 $\epsilon_{ij}$是独立分布的，$\zeta_j \sim N(0,\psi)$ 而 $\epsilon_{ij} \sim N(0,\theta)$。该随机截距模型在计量经济学中通常也被称为"单向误差成分模型(one-way error component model)"。

随机截距或"永久成分"$\zeta_j$ 允许响应的层次在单元之间变动。与固定效应模型相比，这种模型的优势在于，单元间模型不再是饱和的，并且可以纳入时间常数协变量。然而，这些优势也需要付出代价，即依赖于几项假设，如随机截距和协变量之间的零相关假设(参见 3.2.1 节)。

在 $\zeta_j$ 和 $\epsilon_{ij}$是正态的情况下，$\boldsymbol{\beta}$ 的最大似然估计量不能用闭式表示。我们将转而讨论广义最小二乘(generalized least squares, GLS)估计 $\hat{\boldsymbol{\beta}}_{\text{GLS}}$，因为它能够被写成闭式，并且渐近地等价于最大似然估计量。该估计量还具有一个优点，即 $\zeta_j$ 和 $\epsilon_{ij}$的正态性是不需要假设的。

在随机截距模型中，尽管组内估计量(within-estimator)$\hat{\boldsymbol{\beta}}_{\text{W}}$(见 3.6.1 节)是 $\boldsymbol{\beta}$ 的一个无偏的和一致的估计量，但是 GLS 估计量 $\hat{\boldsymbol{\beta}}_{\text{GLS}}$仍是一个最好的线性无偏估计量(BLUE)。GLS 估计量是对组内估计量 $\hat{\boldsymbol{\beta}}_{\text{W}}$和组间估计量(between-estimator)$\hat{\boldsymbol{\beta}}_{\text{B}}$的平均数进行加权而得到的一个矩阵，其中，权重是各个估计量协方差矩阵的反转。在单元间模型

$$\bar{y}_j - \bar{y} = (\bar{x}_j - \bar{x})'\boldsymbol{\beta} + (\bar{\epsilon}_j - \bar{\epsilon}) \tag{3.39}$$

其中，$\boldsymbol{\beta}$ 的组间估计量是 OLS 估计量，即

$$\hat{\boldsymbol{\beta}}_{\text{B}} = \mathbf{B}_{xx}^{-1}\mathbf{B}_{xy}$$

其中，$\mathbf{B}_{xx} \equiv \sum_j (\bar{\mathbf{x}}_j - \bar{\mathbf{x}})(\bar{\mathbf{x}}_j - \bar{\mathbf{x}})'$，$\mathbf{B}_{xy} \equiv \sum_j (\bar{\mathbf{x}}_j - \bar{\mathbf{x}})(\bar{y}_j - \bar{y})$。注意，组间估计量仅仅运用到单元之间的变异，并且忽略来自于纵贯性设计相比较一个截面设计而具有的附加信息。

GLS 估计量(如 Maddala, 1971)于是可表示为

$$\hat{\boldsymbol{\beta}}_{\text{GLS}} = \mathbf{V}_W\hat{\boldsymbol{\beta}}_{\text{W}} + \mathbf{V}_B\hat{\boldsymbol{\beta}}_B$$

其中，权重矩阵为

$$\mathbf{V}_W = (\mathbf{W}_{xx} + \omega^2\mathbf{B}_{xx})^{-1}\mathbf{W}_{xx}$$

$$\mathbf{V}_B = \mathbf{I} - \mathbf{V}_W$$

且

$$\omega = \frac{\theta}{\theta + n\psi}$$

GLS 估计量可用另一方式写为

$$\hat{\boldsymbol{\beta}}_{\text{GLS}} = (\mathbf{W}_{xx} + \omega\mathbf{B}_{xx})^{-1}(\mathbf{W}_{xy} + \omega\mathbf{B}_{xy})$$

从中我们可知，$\omega$ 代表着赋予单元间变异的权重。在固定效应的 OLS 中，$\omega = 0$ 并且变异的来源是被忽略的。对于一个不带有特定单元效应的"单纯"模型，OLS 对应的是 $\omega = 1$，以致所有单元间变异被增加到单元内变异中。通过将特定单元效应看做是随机的，可以在上述处理单元间变异的极端方式之间，提供一种中间方法。还要注意的是，$1 - \omega$相应于将在 7.3.1 节讨论的收缩因子(shrinkage factor)，参见式(7.5)。在对称时点这一特殊情况下，没有缺失数据，且稳定的协变量 $\mathbf{x}_{ij} = \mathbf{x}_i$，得 $\hat{\boldsymbol{\beta}}_{\text{B}} = \mathbf{0}$。由此可知，这种情况下的 GLS 估计量与组间估计量 $\hat{\boldsymbol{\beta}}_{\text{W}}$是相同的。

响应 $\mathbf{y}_j$ 的条件方差，或总残差 $\xi_{ij} = \zeta_j + \epsilon_{ij}$的方差，等于 $\psi + \theta$ 且在时点之间是恒定的。任意两个时点的条件协方差都正好是 $\psi$，并且对应的相关性是以前介绍过的组内相关。

这个**随机截距**协方差结构如展示3.3中A部分所示。注意，此协方差结构是单因子结构的特殊情况，其中，对于所有的 $i,\lambda_i=1$ 且 $\theta_{ii}=\theta$。协方差结构有时被认为是**可交换的**，因为对于某个给定的个人来说，如果残差在时点之间是交叉的，那么残差的联合分布则保持不变。协方差结构也是与球状假设相一致的，即所有成对差异(pairwise differences)的条件方差 $\mathrm{Var}(y_{ij}-y_{i'j}|\mathbf{x}_{ij})$ 都是相等的。注意协变量 $\psi$ 在随机截距模型中都是被约束为非负的。如果这种约束被释放，那么上述协方差结构的构造则经常被称为**复合对称**(compound symmetric)。就对称时点而言，我们可以允许 $\epsilon_{ij}$ 的方差对应每个时点有不同的取值，即 $\theta_{ii}$。

**随机系数模型——增长曲线模型**(Growth curve model)

随机系数模型(如 Laird and Ware,1982)允许响应的层次和协变量的效应在单元间随机地变动。该模型在式(3.12)中被设定为一个双层次模型

$$y_{ij}=\mathbf{x}'_{ij}\boldsymbol{\beta}+\mathbf{z}'_{ij}\boldsymbol{\zeta}_j+\epsilon_{ij}$$

其中，$i=1,2,\cdots,n_j$。在此，$\mathbf{x}_{ij}$ 表示带有固定系数 $\boldsymbol{\beta}$ 的时间变动和时间常数协变量，$z_{ij}$ 表示带有随机系数 $\boldsymbol{\zeta}_j$ 的时间变动协变量。由于随机系数具有零均值，故 $\mathbf{x}_{ij}$ 通常包括 $\mathbf{z}_{ij}$ 中的所有元素，相应的固定效应被解释为平均效应。这些向量中的第一个项通常等于1，相当于一个固定和随机截距。令 $\boldsymbol{\Psi}\equiv\mathrm{Cov}(\boldsymbol{\zeta}_j)$，那么，向量 $\mathbf{y}_j$ 的协方差结构就如展示3.3中的B部分所示。这种特殊情况——残方差被设为在时点之间是相等的，即 $\theta_{ii}=\theta$——是比较普遍的。

纵贯性数据的随机系数模型有一个有用的形式，即增长曲线模型，其中个体不仅被假定为在截距上是不同的，而且其时间轨迹上的其他方面也是不同的，如在响应随着时间的线性增长(或下降)中。这些模型包括关于时间(或时间函数)的随机系数。例如，一个线性增长曲线模型可写为

$$y_{ij}=\mathbf{x}'_{ij}\boldsymbol{\beta}+\zeta_{0j}+\zeta_{1j}t_{ij}+\epsilon_{ij} \tag{3.40}$$

其中，$t_{ij}$(个体 $j$ 在第 $i$ 个时点上的时间)是 $\mathbf{x}_{ij}$ 中的协变量之一。随机截距和斜率不应该被设定为不相关的，因为时间尺度 $t_{ij}$ 的转化形式改变了相关的量值，如3.2.2节中的图3.1所示(也可参见 Elston,1964;Longford,1993)。

这里的协方差结构与式(3.15)中双层次随机系数模型的协方差结构是相同的，明确地展示了式(3.11)中的方差。增长曲线模型的一个路径图如图3.8所示的第一个面板图，其中存在3个时点，分别是 $t_1=0,t_2=1$ 和 $t_3=2$。第二个图表反映的是不对称时点的情况，其中，标有"单元 $j$"的长方形内所有变量都有一个 $j$ 的下标，且在单元间变动。同时，也位于标有"时点 $i$"的长方形内的变量，在时点之间和单元之间都是变动的，且同时具有一个 $i$ 和 $j$ 的下标。从 $t$ 指向 $y$ 的箭头因而代表一个 $y_{ij}$ 在 $t_{ij}$ 上的回归。潜变量 $\zeta_{1j}$ 修正了这个带有 $t_{ij}$ 的回归或交互作用，并因此代表了随机斜率。

就对称时点而言，线性增长曲线模型也可以被公式化为一个双因子模型，即

$$y_{ij}=\lambda_{0i}\eta_{0j}+\lambda_{1i}\eta_{1j}+\epsilon_{ij}$$

其中

$$\eta_{0j}=\beta_0+\zeta_{0j},\quad \eta_{1j}=\beta_1+\zeta_{1j}$$

截距因子 $\eta_{0j}$ 的负荷被固定为 $\lambda_{0i}=1$，斜率因子 $\eta_{1j}$ 的负荷被设为等于 $t_i$。注意，因子均值在此不能像通常在因子模型中的那样被设置为零。

梅雷迪斯和蒂萨克(Meredith and Tisak,1990)提出运用一个双因子模型——类似于图3.8中的第一个图表，但是带有关于 $\eta_{1j}$ 的**自由**因子负荷(服从辨识约束，如 $\lambda_{11}=0$ 和 $\lambda_{12}=1$)——来建构非线性增长模型。传统上，这个因子模型的估计通常要求不带有缺失

数据的对称时点,但这不再是一个约束。然而,如果时点是非常不对称的,在一个已知时点上只有很少的响应,那么就不能再使用因子模型,因为特定时点因子负荷的可靠估计将会受到妨碍。

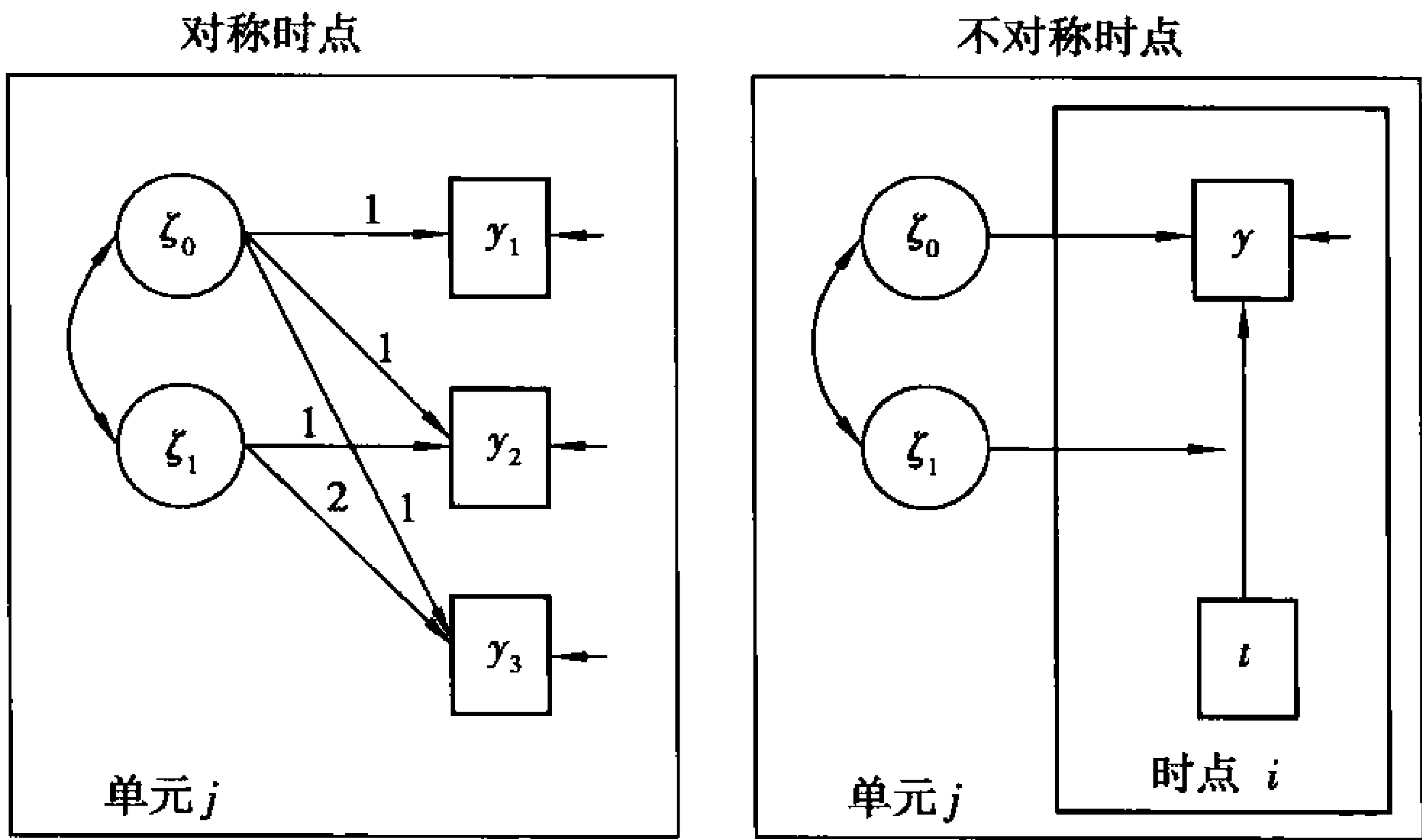



**图 3.8 对称和不对称时点的增长曲线模型路径图**

该模型可以通过运用广义线性混合模型而轻易地扩展到非连续的响应。我们将在 11.3 节估计一个纵贯性计数数据的随机系数模型。

**离散潜变量的纵贯性模型——潜轨迹模型**(latent trajectory model)

有时人们相信,人口是由不同类型或类别的单元——它们是通过随着时间变化的不同发展或发展轨迹的模式而表现出来的——组成的。这些模型都是潜类模型,与上文讨论过的潜在增长模型具有同一形式,除了随机效应现在是离散的。

例如,在一个类似于式(3.40)的线性潜轨迹模型中,类别 $c$ 中某一单元的线性估计量由下式给出,即

$$\nu_{ijc} = e_{0c} + e_{1c}t_{ij}$$

因此,每个潜类别都是由一对系数 $e_{0c}$ 和 $e_{1c}$ 来表现特征的,代表潜轨迹的截距和斜率。对于对称时点,我们不必假设潜轨迹是线性的或者具有另一种特殊形态,但是可为类别 $c(c=1,\cdots,C)$ 设定一个带有潜轨迹

$$\nu_{ijc} = e_{ic}, \quad i = 1,\cdots,I$$

的非结构化模型(unstructured model)。就分类响应而言,潜轨迹模型通常被认为是潜类增长模型(如 Nagin and Land,1993)。它们是混合回归模型在纵贯性数据上的一个应用(如 Quandt,1972)。

如果响应是连续的,这些模型就被认为是潜剖面模型(latent profile model)(如 Gibson,1959),即

$$y_{ijc} = \nu_{ijc} + \epsilon_{ijc}$$

其中,残差 $\epsilon_{ijc}$ 的方差可被允许在类别之间是不同的。潜类模型和潜剖面模型都假设某个单元的响应在给定潜类别隶属时,是条件独立的。穆森和谢登(Muthén and Shedden, 1999)在他们的**增长混合模型**中,通过下述方法释放了关于连续变量的这一假设,即允许残差 $\epsilon_{ijc}$ 是相关的,条件依赖于潜类别隶属,带有一个在类别之间各不相同的协方差矩阵。

## 3.6.2 包含相关残差的模型

随机截距模型包括两个随机项:随机截距和特定时点的残差。虽然随机截距代表随机影响的效应或省略的协变量(随时间变化而保持恒定的),但是残差代表那种即时的、持续时间不超过一个单一时点的随机影响效应。结果产生的复合对称相关结构,通常并

不反映实际中观测到的结果,也就是说,如果响应在发生时间上靠得更近,一对响应之间的(条件)相关倾向于更强。

这种相关结构可以通过允许省略变量的效应是随时间分布的而归纳出来,也导致产生自相关误差(autocorrelated errors)。应当注意,这个省略变量的解释要求影响的总效应(由 $\epsilon_{ij}$表示)在单元上的平均数为零,并且始终与 $x_{ij}$是不相关的(如 Maddala,1977)。在随后的小节中,我们将讨论连续响应的情况,有时也会提及这些模型是如何修正以便适用其他响应类型的。

**自回归残差**

当时点在时间上是等间距的,一个一阶自回归模型 AR(1)可表示为

$$\epsilon_{ij} = \alpha\epsilon_{i-1,j} + \delta_{ij} \tag{3.41}$$

其中,$\epsilon_{i-1,j}$是脱离"创新误差(innovation errors)"$\delta_{ij}$($\delta_{ij} \sim N(0,\sigma_\delta^2)$)而独立分布的。这在图 3.9 中的第一个面板中以路径图的形式进行了说明。注意,在 *AR*(1)模型中,当 $\alpha=1$ 时,可以得到一个"随机游动(random walk)"。

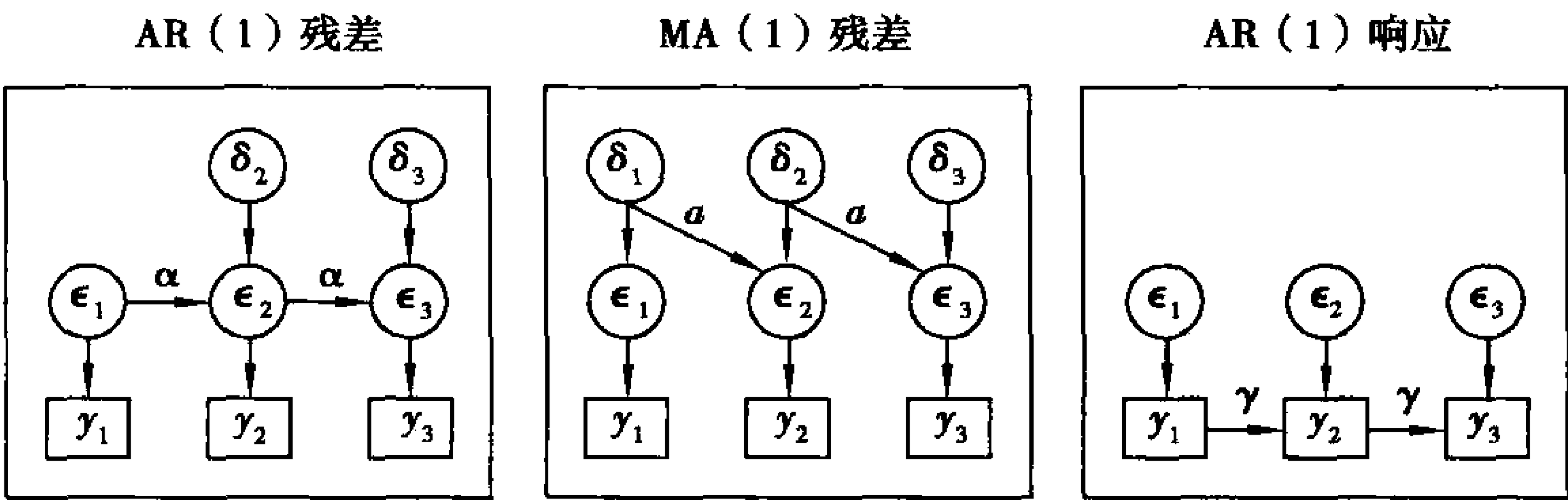



图 3.9 自回归响应以及自回归和移动平均残差的路径图

假定这个过程是**弱平稳的**(weakly stationary),$|\alpha|<1$,那么协方差结构就如 3.6.1 节的展示 3.3 中的 C 部分所示。由此,不同时点上的响应之间的相关性可以被构造为

$$\text{Cor}(\epsilon_{ij},\epsilon_{i+k,j}) = \alpha^k$$

对于非等距的时点,相关结构通常被设定为

$$\text{Cor}(y_{ij},y_{i+k,j}) = \alpha^{|t_{i+k}-t_i|}$$

其中,不对称时点的相关结构是简单地通过用 $t_{ij}$替代 $t_i$而获得的(如 Diggle,1988)。

通常这些一阶自回归协方差结构与复合对称一样,是不符合现实的,因为这些相关性随着时间间隔的渐增而极其迅速地下降。一种可能是去设定一个阶数 $k$ 的高阶自回归过程,即 AR($k$),则

$$\epsilon_{ij} = \alpha_1\epsilon_{i-1,j} + \alpha_2\epsilon_{i-2,j} + \cdots + \alpha_k\epsilon_{i-k,j} + \delta_{ij}$$

另一种可能是将一个随机截距增加到 AR(1)模型中(参见 3.6.5 节的"Hybrid 设定")。就对称时点而言,我们还可以针对每个时点设定一个不同的参数 $\alpha_i$,得到一个残差的**前相关**(antedependence)结构(如 Gabriel,1962)。

**移动平均残差**

随机冲击(random shocks)在其消失之前的某些固定数量的周期中,干扰着响应变量。这样一个过程可以通过移动平均的均值来建立模型(参见 Box et al.,1994)。残差的一个一阶移动平均 MA(1)可以被设定为

$$\epsilon_{ij} = \delta_{ij} + a\delta_{i-1,j}$$

该模型的一个路径图如图 3.9 所示的第二个面板图,其协方差结构展现在 3.6.1 节的展示 3.3 中的 D 部分。我们可知,这个过程"遗忘"了那些在过去不止一个周期发生的

事件,与自回归过程正好相反。

阶数 $k$ 的移动平均模型 MA($k$)被表示为

$$\epsilon_{ij} = \delta_{ij} + a_1\delta_{i-1,j} + a_2\delta_{i-2,j} + \cdots + a_k\delta_{i-k,j}$$

带有延伸到过去 $k$ 个周期的"记忆"。

### 3.6.3 包含延迟响应的模型

在这些模型中,除了 $\mathbf{x}_{ij}$之外,响应 $y_{ij}$的延迟也被纳入作为协变量。对先前响应的依赖被称为"状态依赖(state dependence)";参考 9.6 节的详细说明和一个具体的应用。这些模型也被认为是转移模型(transition model)(如 Diggle et al. ,2002,第 10 章)。当时点在时间上是等间距时,响应 $y_{ij}$的一个一阶自回归模型可被写为

$$y_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \gamma y_{i-1,j} + \epsilon_{ij}$$

假定这个过程是**弱平稳的**,$|\gamma| < 1$,那么协方差结构如 3.6.1 节的展示 3.3 中的 E 部分所示。该模型的一个路径图是图 3.9 中的第三个面板图。

至于残差自回归结构,响应的一阶自回归结构通常被认为是不切实际的,因为这些相关性随着时间间隔的渐增而极其迅速地下降。再一次地,这个问题可以通过设定一个高阶自回归过程 AR($k$)来进行矫正,即

$$y_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \gamma_1 y_{i-1,j} + \gamma_2 y_{i-2,j} + \cdots + \gamma_k y_{i-k,j} + \epsilon_{ij}$$

自回归模型的一个扩展是针对响应 $y_{ij}$的**前相关**模型,它为每个时点设定一个不同的参数 $\gamma_i$。

除了专注于自身的领域,延迟响应模型对于区分不同的纵贯性模型也是非常有用的。现来考虑两个简单模型:一个状态依赖模型——带有一个延迟响应和延迟协变量以及独立的残差 $\epsilon_{ij}$,则

$$y_{ij} = \gamma y_{i-1,j} + \beta_1 x_{ij} + \beta_2 x_{i-1,j} + \epsilon_{ij} \tag{3.42}$$

以及一个不带延迟响应或延迟协变量的自相关模型

$$y_{ij} = \beta x_{ij} + \epsilon_{ij}$$

但是残差 $\epsilon_{ij}$具有一个 AR(1)结构。首先代入式(3.41)$\epsilon_{ij} = \alpha\epsilon_{i-1,j} + \delta_{ij}$,再将 $\epsilon_{i-1,j} = y_{i-1,j} - \beta x_{i-1,j}$代入并重新进行表示,这个自相关模型则可替代性地写为

$$y_{ij} = \alpha y_{i-1,j} + \beta x_{ij} + \alpha\beta x_{i-1,j} + \delta_{ij}$$

注意,这个模型等价于带有约束 $\gamma\beta_1 + \beta_2 = 0$ 的状态依赖模型式(3.42)。重要的是,这意味着我们可以运用状态依赖模型去区别纵贯性模型中的自相关残差和状态依赖。这种真实和"虚假"状态依赖(表面上是状态依赖,而当对残差依赖进行适当模型化时则消失了)之间的区别,在很多应用中都是至关重要的,参考 9.6 节的一个案例。由此还可以得出,在对纵贯性模型建构中的自相关进行一个杜宾-瓦特森检验(Durbin-Watson test)(Durbin and Watson,1950)之前,应当先排除状态依赖模型。否则,一个较大的杜宾-瓦特森统计量——用来说明状态依赖和/或自相关——就是不明确的。

使用延迟响应模型应当谨慎地操作。第一,如果延迟效应不具有一个"因果关系的"解释,那么延迟应当被避免,因为当 $y_{i-1,j}$被纳入作为一个附加协变量时,$\boldsymbol{\beta}$ 的解释则发生了改变。第二,这些模型要求对称的数据,即所有单元都是在相同的时点上被测量的。如果某个单元的响应在一个时点上是缺失的,那么这个完整单元必须被抛弃。第三,延迟响应模型减少了样本量。这是因为在第一个时点上的 $y_{ij}$只能充当协变量,并且不能在延迟响应(它们是缺失的)上进行回归。第四,对于下述这种比较普遍的情况(即当我们

观测某个过程时,它是正在进行的)就出现了一种初始条件问题(如 Heckman,1981b)。

延迟响应模型与自回归残差模型相比,优势在于它们可以便利地应用于连续响应之外的响应类型。

### 3.6.4 其他协方差结构

**无约束模型**

代替建立协方差结构模型的努力,我们可以简单地设定

$$y_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \epsilon_{ij}$$

其中,残差 $\boldsymbol{\epsilon}_j$ 的向量是多元正态的,带有一个无约束的协方差矩阵。这一无约束模型需要的是对称数据,其中 $n_j = I$,并且这个设定相当于一个重复测量多元方差分析(MANOVA)(如 Hand and Crowder,1996,第 2 章)。

该方法为防止某个单元内响应之间相关性的错误设定提供了保障,唯一的假设是,所有单元的响应都是多元正态的,带有相同的残差协方差矩阵。然而,当存在很多时点 $I$ 时,该设定要求较大的样本量,因为有 $I\times(I+1)/2$ 个协方差参数需要和回归系数一起被估计。如果结构形式是有效的,那么无约束模型的估计显然就是没有效率的。

**因子模型**

我们可以通过在线性估计量中纳入因子结构来归纳响应之间的相关性。该方法对于广义线性混合模型来说也是非常有用的,其中,我们通常不能自由设定条件相关(如 Rabe-Hesketh and Skrondal,2001)。

对于一个连续响应,一个关于残差的单因子模型可设定为

$$y_{ij} = \mathbf{x}'_{ij}\beta + \lambda_i\eta_j + \epsilon_{ij}$$

其中,$\mathbf{x}_{ij}$表示带有固定系数$\boldsymbol{\beta}$ 的协变量,$\lambda_i$ 是一个对应时点 $i$ 的因子负荷,$\eta_j$ 表示一个因子。$\xi_{ij} \equiv \lambda_i\eta_j + \epsilon_{ij}$可被视为总残差。注意,该模型的使用需要某种程度的对称,因为每个时点都有一个估计的因子负荷。$\mathbf{y}_j$ 的协方差结构,即一个**因子结构**,如 3.6.1 节展示3.3 中的 F.i. 部分所示。注意,在 $\lambda_i = 1$ 的特殊情况下,上述因子模型就表现为随机截距模型,而当强加一个附加的约束 $\theta_{ii} = \theta$ 时,就得到约束的随机截距模型(产生复合对称性)。对于 3 个时点 $I=3$,单因子模型等价于无约束模型。$I=3$ 的特殊情况还包括一个平稳的一阶自回归残差过程和一个带有一项随机截距的一阶移动平均残差过程(如 Heckman,1981c)。

残差的多维因子模型可被设定为

$$y_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{\lambda}_i\boldsymbol{\eta}_j + \epsilon_{ij}$$

其中,$\mathbf{x}_{ij}$表示带有固定系数$\boldsymbol{\beta}$ 的协变量,$\boldsymbol{\lambda}_i$ 是时点 $i$ 的一个因子负荷向量,$\boldsymbol{\eta}_j$ 表示因子。多维因子结构如 3.6.1 节展示 3.3 中的 F.ii. 部分所示。与测量建模相反,在纵贯性的设置中,我们认为对探索性和验证性因子模型的选择,应当建立在一个现实基础之上,因为因子并没有被附加任何意义。

对于非正态的响应,多维因子模型可以设定为线性估计量的形式,即

$$\nu_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{\lambda}_i\boldsymbol{\eta}_j$$

**混合设定**

我们已经考察过的不同的相关性设定,它们是可以被结合起来的。一个著名的例子是自回归移动平均(ARMA)模型,它是由自回归模型和移动平均模型结合而成。另一种可能是,将随机截距模型与针对响应的一阶自回归过程(如 Jöreskog,1978)或针对残差的

一阶自回归过程(如 Diggle,1988)相结合,从而释放通常在多层次模型中所作出的条件独立性假设。其他方法还包括自回归求和移动平均(ARIMA)模型——运用差分化(differencing)方法来获取平稳性(如 Box et al.,1994)——以及这些模型的特殊形式。

### 3.6.5 非正态响应的广义估计方程

迄今讨论的大多数模型都是基于这样一个概念,即响应之间的相关性(条件依赖于协变量)是可以建模的,并且在某种意义上可以通过潜变量来解释。例如,在增长曲线模型建构中,随机效应抓住了增长轨迹中的个体差异,同时也引发出残差相关。

一种根本不同的方法是,通过运用广义估计方程(GEE),聚焦于均值结构并将相关性归属为一个干扰(如 Liang and Zeger,1986;Zeger and Liang,1986),参见6.9 节。最简单的方式是去估计均值结构,权当响应都是独立的,然后运用所谓的夹挤估计量(sandwich estimator)来修正相关性的标准误差(参见 8.3.3 节)。参数估计可以被证明为是前后一致的,但如果响应是相关的,它们就会是无效的。为了增加效果,故要在一个针对广义线性模型的迭代重加权最小二乘算法(iteratively reweighted least squares algorithm)的多元扩展内,设定一个"工作相关矩阵(working correlation matrix)"(参见 6.9 节的详细讨论)。通常,展示3.3 中列出的那些结构,其中之一被用来表示残差 $y_{ij}-g^{-1}(\mathbf{x}_{ij}'\boldsymbol{\beta})$ 的工作相关矩阵,以及无约束结构和独立相关结构的工作相关矩阵。工作相关矩阵是与适当广义线性模型的方差函数结合在一起的,如果响应是计数的话,通常需要考虑到过度离散。必须注意的是,除了连续响应之外,设定的相关结构一般不能从统计模型中衍生出来。因此,其中不存在似然,GEE 是一种多元准似然方法(multivariate quasi-likelihood approach)。

一般而言,用 GEE 进行估计的回归系数与包含潜变量的模型的回归系数相比,具有一个不同的解释。后者代表已知潜变量的协变量**条件效应**,在纵贯性设置中代表的是**特定单元效应**。另一方面,GEE 则提供**边际的**或**总体平均**效应,这里,个体的差异是平均的,而不是由潜变量来建模。在正态概率和对数概率回归中,边际效应与条件效应相比,是倾向于衰减的(趋于0),如 1.4 节图 1.6 所示的正态概率回归。对于其他链接和含有随机系数的模型(除去含有同一性链接的模型以及含有一个对数链接和一个正态分布随机截距的模型),边际效应与条件效应之间也会出现差异(同时参考 4.8.1 节)。

注意,也存在带有相应似然的"适当(proper)"边际统计模型。实例包括巴哈杜尔模型(Bahadur,1961)(它经由边际相关来对相关性进行参数化)和达勒模型(Dale,1986)(它经由边际二元优比(odds-ratios)对相关性进行参数化);参见费兹马里斯等人(Fitzmaurice et al.,1993)和莫伦博格斯(Molenberghs,2002)的介绍。

条件效应或边际效应能否引起关注,将取决于具体情境。例如,在公共卫生方面,总体平均效应可能是受关注的,而条件效应对于患者和临床医师来说则是非常重要的。重要的是,通过整合潜变量,边际效应可以从条件模型中衍生出来。遗憾的是,条件效应一般不能从边际效应中衍生出来。条件效应在总体之间可能是对称的。然而,如果条件效应在两个总体中是相同的,但是随机截距方差不同,那么边际效应也将是不同的。

注意,希格蒂和泽戈(Heagerty and Zeger,2000)介绍了潜变量模型,其中,边际均值如同 GEE 中一样在协变量上进行回归。在这些模型中,条件均值(已知潜变量)和协变量之间的关系是通过解答连接条件和边际均值的积分方程(参见 4.8.1 节的方程(4.28))而建立起来的。有趣的是,涉及的这个积分可被写为模型中随机部分各项之和的分布上的一维积分。

## 3.7 总结与进一步阅读

我们已经描述了经典潜变量模型,诸如多层次回归模型、测量模型、探索性和验证性因子模型、题器响应模型、结构方程模型、潜类模型和纵贯性数据的几个模型。一个关于这些经典潜变量模型的统一框架(将它们与第 2 章所介绍的响应过程结合在一起)将在下一章进行阐述。有些经典潜变量模型也会运用于本书的应用部分(特别是第 9 章),尽管大多数应用都是基于扩展模型的。

多层次模型的相关书籍很多,如(近似地以难度排序)Kreft and de Leeuw(1998)、Hox(2002)、Raudenbush and Bryk(2002)、Snijders and Bosker(1999)、Aitkin et al.(2004)、Goldstein(2003)、Longford(1993)、Cox and Solomon(2002)、McCulloch and Searle(2001)。

我们仅仅展示了测量理论的冰山一角,而没有进行深入探讨(如普遍性的理论)(参见 Cronbach et al.,1972;Shavelson and Webb,1991;Brennan,2001)。作为介绍性的阅读,我们推荐 Streiner and Norman(1995)。中等层次的处理包括 Crocker and Algina(1986)、Dunn(2004)。劳德和诺维卡提供了一个更高级和权威的处理(Lord and Novick,1968)。劳利和马克斯维尔(Lawley and Maxwell,1971)以及穆拉克(Mulaik,1972)的著作对于连续响应的因子模型非常有用,而巴托罗买和诺特(Bartholomew and Knott,1999)也考虑了二项、多项以及混合响应。

题器响应理论方面的书籍包括 Lord and Novick(1968)、Lord(1980)、Hambleton and Swaminathan(1985)、Hambleton et al.(1991)、van der Linden and Hambleton(1997)、Embretson and Reise(2000)、De Boeck and Wilson(2004)。我们没有讨论非参数题器响应理论(如 Sijtsma and Molenaar,2002)或展开(理想点)模型,其中题器特质曲线是非单调的(如 Coombs,1964;Roberts et al.,2000)。

麦克彻恩(McCutcheon,1987)对潜类模型进行了介绍,而克劳格(Clogg,1995)给出了一项调查研究。关于混合模型的书籍包括 Everitt and Hand(1981)、McLachlan and Peel(2000)、Böhning(2000)。

有关结构方程模型的参考书目有 Dunn et al.(1993)、Bollen(1989)、Wansbeek and Meijer(2002)。

我们没有讨论纵贯性数据的状态空间(state-space)模型(如 Jones,1993)或隐藏的马尔可夫(潜在转化)模型(如 MacDonald and Zucchini,1997;van de Pol and Langeheine,1990)。关于纵贯性数据建模的有用书籍包括 Hand and Crowder(1996)、Crowder and Hand(1990)、Hsiao(2002)、Baltagi(2001)、Diggle et al.(2002)、Lindsey(1999)、Everitt and Pickles(1999)。

# 4 一般模型框架

## 4.1 介　绍

一般模型框架归纳和概括了多层次模型、因子模型、题器响应模型、潜类模型及结构方程模型，以及在第2章中讨论过的纵贯性模型。

在第2章中，我们主要关注了含连续响应的模型。现在介绍一下能够适用于在第2章讨论过的所有响应过程的潜变量模型。与第2章中的模型不同的是，随机系数和因子被包括在同一个模型中，潜变量也可以在多个层次上变化，得到诸如多层次因子模型之类的模型。可设定多层次结构方程，使其对在同一层次或者更高层次的潜变量或观测变量回归。使用其他的连续分布或离散分布，或者非参数最大似然估计法，我们能够释放潜变量的多元正态分布假设，同时，也能适用于不同类型的潜类模型。模型框架主要与拉贝-赫斯基思等(Rabe-Hesketh et al.,2004a)阐述的广义线性潜混合模型(Generalized Linear Latent And Mixed Model,缩写为GLLAMM)一致，也可参考拉贝-赫斯基思等(Rabe-Hesketh et al.,2001a)。除此之外，本章还要论述在这类框架中未能整合进去的模型，如多层次潜类模型。

一般模型方程的实质是分级条件关系的设定：响应模型设定观测响应的分布，与潜变量和协变量(通过一个线性估计量和链接函数)之间是条件依赖关系，而且在结构模型中，潜变量自身也能对其他潜变量和观测到的协变量回归。最终，结构模型中随机干扰项的分布得以设定。综上所述，本章的4.2节至4.3节主要关注以下内容：

- 响应模型。
- 潜变量的结构模型。
- 结构模型中干扰项的分布。

在模型设定中，较为关键的部分涉及在模型参数中进行必要约束的问题。因此，这里介绍了不同类型的参数约束，同时，基本参数的相关概念也将在4.5节中进行介绍。

为了能完全理解一个潜变量模型，很重要的一点是，要考察响应变量的矩结构，它边缘于潜变量，但条件依赖于观测协变量。为了推导这一点，可从4.6节所提到的潜变量以及线性估计量的简化式开始推导。然后，在4.7节中，通过积出结构模型的干扰项得到潜变量矩结构(此节和前一节有些偏重技术性，如果不需要则可忽略)。得到矩结构后，在4.8节，我们再通过积出潜变量，导出响应的矩结构。这有助于更好地认识在潜变量模型中，条件协变量效应与边缘协变量效应之间的区别。最后，在4.9节，得出简化式

分布以及观测响应的条件分布(假设是解释变量),这也是似然的基础。简化式参数的概念对于第5章中的辨识和等价的介绍很重要,将在4.10节中论及。

## 4.2 响应模型

响应模型条件依赖于潜变量,可设定一个包含线性估计量$\nu_i$,链接函数$g(\cdot)$以及指数族分布的广义线性模型

$$f(y_i \mid \theta_i, \phi) = \exp\left\{\frac{y_i\theta_i - b(\theta_i)}{\phi} + c(y_i, \phi)\right\}$$

其中,$\theta_i$是均值$\mu_i = g^{-1}(\nu_i)$的函数,$\nu_i$取决于潜变量。对于响应变量,可设定任意广义线性模型的条件密度,包括在第2章中所介绍的扩展部分。另外,也可设定尺度参数模型和阈值模型。表4.1列举了可操作的响应类型和对每种类型进行介绍、应用的章节。

**表4.1 涉及的响应类型和相应应用章节**

| 响应类型 | 章　节 |
|---|---|
| 连续性 | |
| 二项 | 第9章 |
| 定序 | 第10章 |
| 事件数 | 第11章 |
| 时间存续 | 第12章 |
| 　离散时间存续 | |
| 　连续时间存续 | |
| 对照 | 第13章 |
| 　定类变量 | |
| 　排序 | |
| 　成对比较 | |
| 混合响应 | 第14章 |

在4.2.1节中,我们将传统的随机系数和因子模型结合起来,从而得到了在4.2.2节中所介绍的一般模型的"广义因子(generalized factor,GF)"方程。这个广义因子方程是**多元的**;矩阵表达式设定了多元响应线性估计量的向量。这个多元方程对于推导观测响应(连续个案)或潜响应(二项、定序或对照案例)的协方差结构是很有帮助的。

采用在4.2.3节中论述的"广义随机系数(generalized random coefficient,GRC)"方程,可以对线性估计量进行界定。这个**单变量的**方程,类似于在第3章中讲过的多层次随机系数模型的单变量方程。GRC方程的一个重要的优势在于,它含有参数和协变量的独立项,从而与GF方程比,使得模型结构更为清晰。在4.2.4节,GF和GRC方程同时被用来设定一个双层次因子模型,通过4.2.5中的两个方程,还能对探索性的潜类模型进行设定。

### 4.2.1 结合传统随机系数模型和因子模型

在3.2节和3.3节论述过的传统随机系数和因子模型，两者比人们通常认识到的更相似。首先从下面的等式中回顾一下随机系数模型(3.14)

$$\mathbf{y}_j = \mathbf{X}_j\boldsymbol{\beta} + \mathbf{Z}_j\boldsymbol{\eta}_j + \boldsymbol{\epsilon}_j \tag{4.1}$$

以及(3.34)节中结构等式模型的测量部分

$$\mathbf{y}_j = (\boldsymbol{\nu} + \mathbf{K}\mathbf{x}_j) + \boldsymbol{\Lambda}\eta_j + \boldsymbol{\epsilon}_j \tag{4.2}$$

在这两个等式中，为了简化起见，很多上下标都省略了。

尽管在解释上有所不同，这些模型都有相似的结构。在随机系数模型中，$\mathbf{y}_j$ 表示一层次单元响应值的向量，在第 $j$ 个二层次单元内，而 $\mathbf{y}_j$ 表示因子模型中的题器。无论单元在多级设置的单元中或者因子模型的题器中，是否为最低一级，为了便于进一步发展，都将初始单元 $i$ 视为一层次单元。而随机系数模型中的群集 $j$，或因子模型中的其他单元则成为二层次单元。随机系数模型中的干扰项 $\boldsymbol{\epsilon}_j$ 相当于因子模型中的唯一因子，即"误差"。随机系数模型中的随机效应 $\eta_j$ 相当于因子模型中的公共因子。在随机效应或者公共因子 $\eta_j$ 的处理上，可以采用**潜变量**进行分析。

随机效应的设计矩阵 $\mathbf{Z}_j$ 的因子负荷与矩阵 $\boldsymbol{\Lambda}$ 相当。矩阵 $\mathbf{Z}_j$ 和矩阵 $\boldsymbol{\Lambda}$ 有两点不同，首先，矩阵 $\mathbf{Z}_j$ 是一个由协变量和常数组成的已知矩阵，而后者是一个未知参数矩阵。再者，$\mathbf{Z}_j$ 能够在二层次单元 $j$ 中变动，而 $\boldsymbol{\Lambda}$ 是恒量。然而，这里要将 $\mathbf{Z}_j$ 和 $\boldsymbol{\Lambda}$ 都作为**结构矩阵**，用 $\boldsymbol{\Lambda}_j$ 表示。

这两个模型中的固定部分 $\mathbf{X}_j\boldsymbol{\beta}$ 和 $(\boldsymbol{\nu}+\mathbf{K}\mathbf{x}_j)$ 可来设定同一个均值结构。在单共变函数的情况下，第 $i$ 行或者第 $i$ 个一层次单元的各项分别是 $\beta_0 + x_{ij}\beta$ 和 $v_i + k_i x_j$。然而，前者假设了一个一层次特定协变量 $x_{ij}$ 的常数效应 $\beta$，后者假设了一个二层次特定协变量 $x_j$ 的在第一层次的特定效应 $k_i$。即便如此，这个差异还只是停留在表面上：在一个随机系数模型中，与第一层次单元 $i$ 的虚拟变量的交互作用，可用来使其系数取决于 $i$ 值；在因子模型中，不同的协变量能够使不同的 $i$ 值表示一个一层次的特定单元的协变量。

整合概括随机系数模型和因子模型的**响应模型**可表达为

$$\mathbf{y}_j = \mathbf{X}_j\boldsymbol{\beta} + \boldsymbol{\Lambda}_j\boldsymbol{\eta}_j + \boldsymbol{\epsilon}_j \tag{4.3}$$

其中，结构矩阵 $\boldsymbol{\Lambda}_j$ 包含变量和参数。下面的展示4.1是符号和各术语的汇总：

展示4.1 符号和术语的汇总

| 联合模型 | | 随机系数模型 | | 因子模型 | |
|---|---|---|---|---|---|
| 符号 | 术语 | 符号 | 解释 | 符号 | 解释 |
| $i$ | 一层次单元 | $i$ | 一层次单元 | $i$ | 题器 |
| $j$ | 二层次单元 | $j$ | 二层次单元 | $j$ | 单元 |
| $\mathbf{y}_j$ | 响应 | $\mathbf{y}_j$ | 响应 | $\mathbf{y}_j$ | 响应 |
| $\boldsymbol{\epsilon}_j$ | 误差 | $\boldsymbol{\epsilon}_j$ | 残差项 | $\boldsymbol{\epsilon}_j$ | 唯一因子 |
| $\boldsymbol{\eta}_j$ | 潜变量 | $\boldsymbol{\eta}_j$ | 随机系数 | $\boldsymbol{\eta}_j$ | 公共因子 |
| $\boldsymbol{\Lambda}_j$ | 结构矩阵 | $\mathbf{Z}_j$ | 设计矩阵 | $\boldsymbol{\Lambda}$ | 因子负荷矩阵 |
| $\mathbf{X}_j\boldsymbol{\beta}$ | 固定部分 | $\mathbf{X}_j\boldsymbol{\beta}$ | 固定部分 | $(\boldsymbol{\nu}+\mathbf{K}\mathbf{x}_j)$ | 固定部分 |

如果潜变量是离散的，这个模型框架能用来形成不同类型的潜类模型。一个混合回归模型（如潜类增长模型）（参见 3.3 节）通过应用一个随机系数模型中的离散潜变量就能轻易得到；可参见 9.5 节，在元分析中的例子。4.2.5 节讲述了如何应用这个框架形成探索性潜类模型（exploratory latent class models）。

要注意的是，将一个因子模型的题器或者包含任何多元响应的变量视为一层次单元，并且将初始单元处理为二层次群集，这是在多元、多层次的回归建模中是一个普遍的方法，如（Goldstein，2003）第 6 章。这种方法的一个优势在于，如果响应随机性缺失（MAR）（见 8.3.1 节，缺失类型），缺失的响应仅仅改变群集的大小，这可通过多层次建模软件进行处理。劳登布什和桑普森（Raudenbush and Sampson，1999ab），劳登布什和布莱克（Raudenbush and Bryk，2002）在单参数题器响应模型的论述中，以及德伯克和威尔逊（De Boeck and Wilson，2004）在双参数题器响应模型中，也采用过这种方法。

### 4.2.2 广义因子（GF）方程中的线性估计量

考察线性估计量的主要优势在于，我们在第 2 章中提到的所有的响应进程都能进行调适。

这里，可将式（4.3）中的联合模型表达成广义因子（GF）方程，并为 $\mathbf{y}_j$ 设定一个同一性链接和正态密度，$j$ 单元响应的线性估计量的向量表示为

$$\boldsymbol{\nu}_j = \mathbf{X}_j\boldsymbol{\beta} + \boldsymbol{\Lambda}_j\boldsymbol{\eta}_j$$

在介绍拓展到多层次之前，我们将重新介绍在 3.2.1 节中使用过的模型层次的上下标，并将模型表达为

$$\boldsymbol{\nu}_{j(2)} = \mathbf{X}_{j(2)}\boldsymbol{\beta} + \boldsymbol{\Lambda}_{j(2)}\boldsymbol{\eta}_j^{(2)} \tag{4.4}$$

以 $j(2)$ 为下标的向量，包含**第 $j$ 个第二层次单元的所有要素**，而（2）为上标的潜变量**在第二层次变动**。要注意的是，对任意双层次模型来说，$\boldsymbol{\eta}_{j(2)} = \boldsymbol{\eta}_j^{(2)}$。

在 4.2.2 节展示 4.2A 中，用 GF 方程来表示一个有 5 个题器的一层次双因子模型的随机部分，在这个模型中，前 3 个题器负荷于因子 1，后两个题器负荷于因子 2。4.2.2节展示 4.3A 使用了相同的符号来表示一个双层次随机系数模型。这里，在第 $j$ 个二层次单元上共有 3 个一层次单元，同时这个模型包括了协变量 $t_{ij}$ 的一个随机截距和随机斜率。

现在，可将该模型推广到 $L$ 层次，其结果为

$$\boldsymbol{\nu}_{z(L)} = \mathbf{X}_{z(L)}\boldsymbol{\beta} + \sum_{l=2}^{L}\boldsymbol{\Lambda}_{z(L)}^{(l)}\boldsymbol{\eta}_{z(L)}^{(l)} \tag{4.5}$$

其中，$\boldsymbol{\nu}_{z(L)}$ **是特定层次 $L$（最高层次）中的所有单元的线性估计量的向量，单元 $z$ 和** $\eta_{z(L)}^{(l)}$ **为该 $L$ 层单元**（参见 4.2.3 节的展示 4.4 的右下图）**的所有处于 $l$ 层的潜变量的向量。**$\boldsymbol{\Lambda}_{z(L)}^{(l)}$ 的上标（$l$）表示矩阵是 $l$ 层次潜变量的特定矩阵。正如第 3 章的展示 3.2 所示，我们也可将该模型改写为下面这样的双层次模型：

$$\boldsymbol{\nu}_{z(L)} = \mathbf{X}_{z(L)}\boldsymbol{\beta} + \boldsymbol{\Lambda}_{z(L)}\boldsymbol{\eta}_{z(L)} \tag{4.6}$$

其中，$\boldsymbol{\Lambda}_{z(L)} = [\boldsymbol{\Lambda}_{z(L)}^{(2)}, \cdots, \boldsymbol{\Lambda}_{z(L)}^{(L)}]$，并且 $\boldsymbol{\eta}_{z(L)} = (\boldsymbol{\eta}_{z(L)}^{(2)\prime}, \cdots, \boldsymbol{\eta}_{z(L)}^{(L)\prime})'$ 是 $L$ 层的第 $z$ 个单元的所有潜变量的向量（见展示 4.4 的右上图）。

展示 4.2　单层次双因子模型的随机部分

A. GF 方程：

$$\boldsymbol{\Lambda}_{j(2)}\boldsymbol{\eta}_j^{(2)} = \begin{bmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & \lambda_{42} \\ 0 & \lambda_{52} \end{bmatrix} \begin{bmatrix} \eta_{1j}^{(2)} \\ \eta_{2j}^{(2)} \end{bmatrix} = \begin{bmatrix} \lambda_{11}\eta_{1j}^{(2)} \\ \lambda_{21}\eta_{1j}^{(2)} \\ \lambda_{31}\eta_{1j}^{(2)} \\ \lambda_{42}\eta_{2j}^{(2)} \\ \lambda_{52}\eta_{2j}^{(2)} \end{bmatrix}$$

B. GRC 方程的矩阵式：

$$\sum_{m=1}^{M}\eta_{mj}^{(2)}\mathbf{Z}_{mj}^{(2)}\boldsymbol{\lambda}_m^{(2)} = \eta_{1j}^{(2)} \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{\mathbf{Z}_{1j}^{(2)}} \underbrace{\begin{bmatrix} \lambda_{11} \\ \lambda_{21} \\ \lambda_{31} \end{bmatrix}}_{\boldsymbol{\lambda}_1^{(2)}} + \eta_{2j}^{(2)} \underbrace{\begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}}_{\mathbf{Z}_{2j}^{(2)}} \underbrace{\begin{bmatrix} \lambda_{42} \\ \lambda_{52} \end{bmatrix}}_{\boldsymbol{\lambda}_2^{(2)}}$$

$$= \begin{bmatrix} \lambda_{11}\eta_{1j}^{(2)} \\ \lambda_{21}\eta_{1j}^{(2)} \\ \lambda_{31}\eta_{1j}^{(2)} \\ \lambda_{42}\eta_{2j}^{(2)} \\ \lambda_{52}\eta_{2j}^{(2)} \end{bmatrix}$$

C. GRC 方程：

$$\sum_{m=1}^{M}\eta_{mj}^{(2)}\mathbf{z}_{mij}^{(2)\prime}\boldsymbol{\lambda}_m^{(2)} = \eta_{1j}^{(2)}\boldsymbol{\delta}'_{1i}\boldsymbol{\lambda}_1^{(2)} + \eta_{2j}^{(2)}\boldsymbol{\delta}'_{2i}\boldsymbol{\lambda}_2^{(2)}$$

其中，$\delta'_{1i}$是 $\mathbf{Z}_{1j}^{(2)}$ 的第 $i$ 行；$\delta'_{2i}$是 $\mathbf{Z}_{2j}^{(2)}$ 的第 $i$ 行。

当 $i=1$ 时：

$$\eta_{1j}^{(2)}\boldsymbol{\delta}'_{1i}\boldsymbol{\lambda}_1^{(2)} + \eta_{2j}^{(2)}\boldsymbol{\delta}'_{2i}\boldsymbol{\lambda}_2^{(2)} = \eta_{1j}^{(2)}[100]\begin{bmatrix} \lambda_{11} \\ \lambda_{21} \\ \lambda_{31} \end{bmatrix} + \eta_{2j}^{(2)}[00]\begin{bmatrix} \lambda_{42} \\ \lambda_{52} \end{bmatrix} = \lambda_{11}\eta_{1j}^{(2)}$$

展示 4.3　双层次随机系数模型的随机部分

A. GF 方程：

$$\boldsymbol{\Lambda}_{j(2)}\boldsymbol{\eta}_j^{(2)} = \begin{bmatrix} 1 & t_{1j} \\ 1 & t_{2j} \\ 1 & t_{3j} \end{bmatrix} \begin{bmatrix} \eta_{1j}^{(2)} \\ \eta_{2j}^{(2)} \end{bmatrix} = \begin{bmatrix} \eta_{1j}^{(2)} + t_{1j}\eta_{2j}^{(2)} \\ \eta_{1j}^{(2)} + t_{2j}\eta_{2j}^{(2)} \\ \eta_{1j}^{(2)} + t_{3j}\eta_{2j}^{(2)} \end{bmatrix}$$

B. GRC 方程矩阵式：

$$\sum_{m=1}^{M}\eta_{mj}^{(2)}\mathbf{Z}_{mj}^{(2)}\boldsymbol{\lambda}_m^{(2)} = \eta_{1j}^{(2)} \underbrace{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}}_{\mathbf{Z}_{1j}^{(2)}} \underbrace{[1]}_{\lambda_1^{(2)}} + \eta_{2j}^{(2)} \underbrace{\begin{bmatrix} t_{1j} \\ t_{2j} \\ t_{3j} \end{bmatrix}}_{\mathbf{Z}_{2j}^{(2)}} \underbrace{[1]}_{\lambda_2^{(2)}} = \begin{bmatrix} \eta_{1j}^{(2)} + t_{1j}\eta_{2j}^{(2)} \\ \eta_{1j}^{(2)} + t_{2j}\eta_{2j}^{(2)} \\ \eta_{1j}^{(2)} + t_{3j}\eta_{2j}^{(2)} \end{bmatrix}$$

C. GRC 方程：

$$\sum_{m=1}^{M}\eta_{mj}^{(2)}\mathbf{z}_{mij}^{(2)\prime}\boldsymbol{\lambda}_m^{(2)} = \eta_{1j}^{(2)} \times 1 \times 1 + \eta_{2j}^{(2)}t_{ij} \times 1 = \eta_{1j}^{(2)} + \eta_{2j}^{(2)}t_{ij}$$

### 4.2.3 广义随机系数(GRC)方程的线性估计量

为了简单起见，首先考察一个双层次模型。在 GF 方程(4.4)中，结构矩阵 $\Lambda_{j(2)}$ 既不是一个纯粹的设计矩阵，也不是一个纯粹的参数矩阵。相反，它同时含有变量和参数。这里可以根据纯粹的设计矩阵 $\mathbf{Z}_{mj}^{(2)}$ 和纯粹的参数向量 $\boldsymbol{\lambda}_m^{(2)}$ 来展开矩阵 $\Lambda_{j(2)}\boldsymbol{\eta}_j^{(2)}$，这样式(4.4)则变为

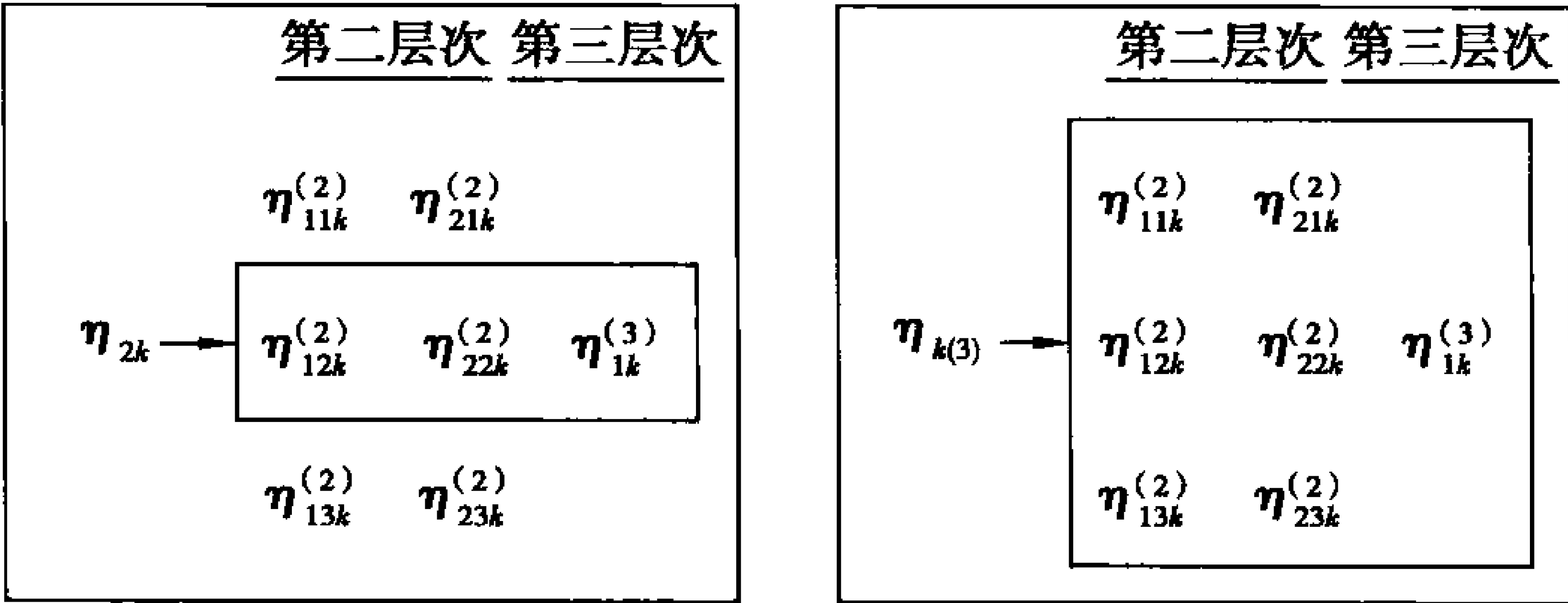
$$\boldsymbol{\nu}_{j(2)} = \mathbf{X}_{j(2)}\boldsymbol{\beta} + \underbrace{\sum_{m=1}^{M}\eta_{mj}^{(2)}\mathbf{Z}_{mj}^{(2)}\boldsymbol{\lambda}_m^{(2)}}_{\Lambda_{j(2)}\boldsymbol{\eta}_j^{(2)}} \tag{4.7}$$

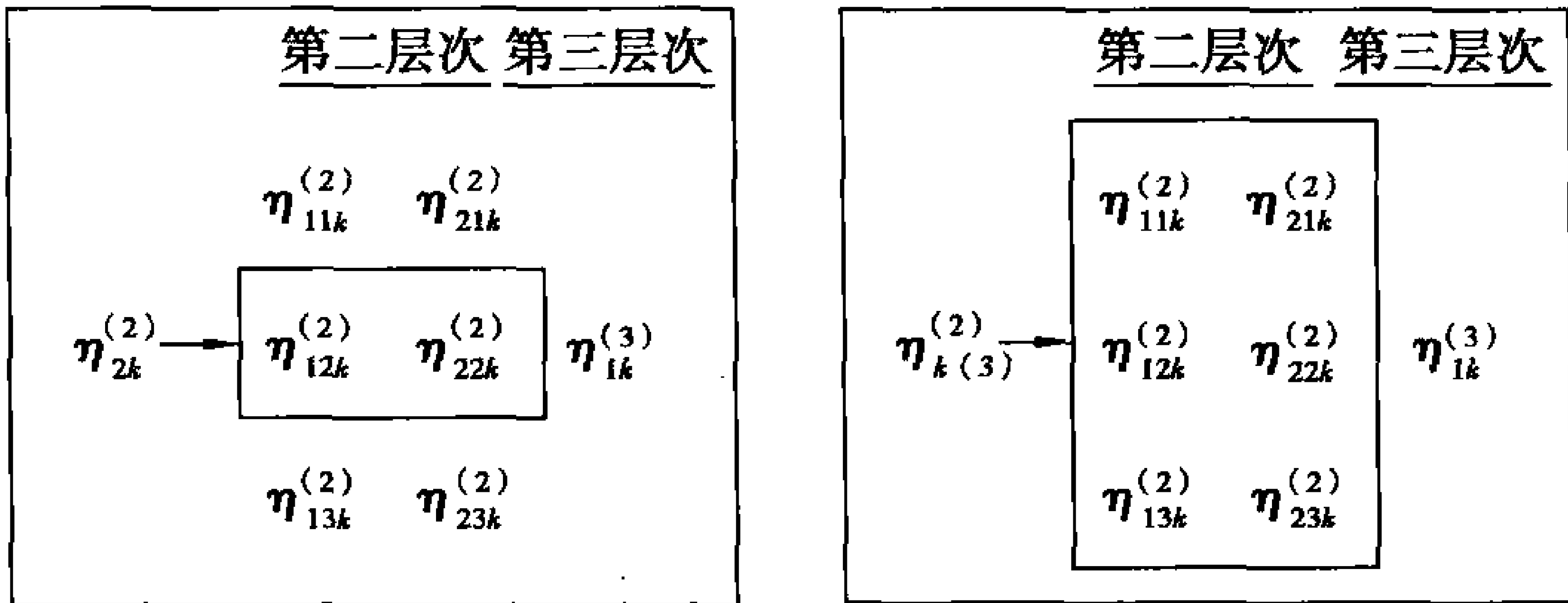
其中，$\eta_{mj}^{(2)}$ 是第 $m$ 个潜变量，$\mathbf{Z}_{mj}^{(2)}$ 是一个包含协变量和固定已知常数的 $n_j^{(2)} \times p_m^{(2)}$ 的(设计)矩阵，$\boldsymbol{\lambda}_m^{(2)}$ 是与第 $m$ 个潜变量关联的含 $p_m^{(2)}$ 个参数的向量。乘积 $\mathbf{Z}_{mj}^{(2)}\boldsymbol{\lambda}_m^{(2)}$ 表示 $\Lambda_{j(2)}$ 的第 $m$ 列；因而 $\boldsymbol{\lambda}_m^{(2)}$ 不是矩阵 $\Lambda_{j(2)}$ 中的向量。我们将这个方程称为“GRC 方程的矩阵式”。

如果在相应的矩阵 $\mathbf{Z}_{mj}^{(2)}$ 中的所有元素均为 0 或 1，通常潜变量 $\eta_{mj}^{(2)}$ 就能被解释为一个**因子**。在这种情形下，$\boldsymbol{\lambda}_m^{(2)}$ 包含该因子的 $p_m^{(2)}$ 个非零因子负荷，而 $\mathbf{Z}_{mj}^{(2)}$ 的作用是给不同的题器赋予正确的因子负荷。这在展示 4.2B 中可以看到，其中，$\mathbf{Z}_{mj}^{(2)}$ 的各列是表示各个题器在第 $m$ 个因子上是否有负荷的虚拟变量，即除了那些有相关题器的对应位置，向量的其他值都等于 0。如果把所有题器都当做第一层次的单元，以及把所有的响应都集中在单独一个响应向量 $\mathbf{y}_{j(2)}$ 中，那么因子模型的这种方程表达方式是很自然的，因为在这种情形下，采用虚拟变量是“指涉”某个题器简单明了的方法。在展示 4.2B 的例子中，“GRC 方程的矩阵式”可简单地理解为“公共因子 1 由题器 1 到 3 所测量，其因子负荷为 $\boldsymbol{\lambda}_1^{(2)}$”，同时，“公共因子 2 由题器 4 到 5 所测量，其因子负荷为 $\boldsymbol{\lambda}_2^{(2)}$”。需要注意的是，在展示 4.2 中，为了与传统因子模型的标记保持一致，$\boldsymbol{\lambda}_m^{(2)}$ 的元素索引被表示为 $\lambda_{rm}$，但在本书的剩余章节中，$\boldsymbol{\lambda}_m^{(2)}$ 中的第 $r$ 个元素将被表示成 $\lambda_{mr}^{(2)}$。

展示 4.4 不同类型的潜变量向量

考察一个处于第三层次的单元 $k$，它包括 3 个第二层次单元。在第三层次有一个潜变量($M_3 = 1$)，在第二层次有两个潜变量($M_2 = 2$)。其元素为 $\eta_{mjk}^{(2)}, m = 1,2, j = 1,2,3$ 以及 $\eta_{mk}^{(3)}, m = 1$。






---

如果对应的矩阵 $\mathbf{Z}_{mj}^{(2)}$ 是一个列向量，且其对应的标量"因子负荷"$\lambda_m^{(2)}$ 被设定为 1，那么潜变量 $\eta_{mj}^{(2)}$ 就是**某个变量** $z_{ij}$ **的随机系数**，如展示 4.3 所示。

式(4.7)的第 $i$ 行则变为

$$\nu_{ij} = \mathbf{x}_{ij}'\boldsymbol{\beta} + \sum_{m=1}^{M} \eta_{mj}^{(2)} \mathbf{z}_{mij}^{(2)\prime} \boldsymbol{\lambda}_m^{(2)} \tag{4.8}$$

其中，$\mathbf{x}_{ij}'$ 是 $\mathbf{X}_{j(2)}$ 的第 $i$ 行，$\mathbf{z}_{mij}^{(2)\prime}$ 是 $\mathbf{z}_{mj}^{(2)}$ 的第 $i$ 行。这即为双层次模型的 GRC 方程式。

正如在 4.2.2 节的展示 4.2C 中看到的那样，要用这个符号表现出因子模型，则需要含 $p_m^{(2)}$ 个元素（其中 $p_m^{(2)}$ 是用来测量或者在第 $m$ 个因子有"负荷"的题器的个数）的虚拟向量 $\boldsymbol{\delta}_{mi}$，如果题器 $i$ 于因子 $m$ 上有负荷，则 $\boldsymbol{\lambda}_m^{(2)}$ 的该元素等于 1，否则为 0。展示 4.3C 显示了用这种标记所表示的随机系数模型。这里，$\mathbf{z}_{mij}^{(2)}$ 是标量，其对应的参数 $\lambda_m = 1$。

这个模型可被扩展到 $L$ 层次，公式为

$$\nu_i = \mathbf{x}_i'\boldsymbol{\beta} + \sum_{l=2}^{L} \sum_{m=1}^{M_l} \eta_m^{(l)} \mathbf{z}_{mi}^{(l)\prime} \boldsymbol{\lambda}_m^{(l)} \tag{4.9}$$

其中，$M_l$ 是第 $l$ 层次上的潜变量的数目，为了简化表达，省略了较高层次上的观测标记。

要了解更多的 GRC 方程信息可参考 Rabe-Hesketh and Pickles(1999)、Rabe-Hesketh et al. (2000)。

## 4.2.4 例：GF 和 GRC 方程中的双层次因子模型

现在，我们已能定义一个双层次因子模型，如果给题器提供响应的单元套嵌于群集之中，这样一个模型会有一定用处，如关于学校的学生的研究。在这样的案例中，单层次因子模型并不适宜，因为来自于同一个群集中的不同单元的响应，很可能是相关的。对于连续响应或者潜在响应，通常需要假设：

$$\mathbf{y}_{j(2)}^* \sim \mathrm{N}(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_1)$$
$$\boldsymbol{\mu}_j \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}_2)$$

其中，$\boldsymbol{\mu}_j$ 和 $\boldsymbol{\mu}$ 是截距的向量。然后分别在单元和群集的层次上分别设定公共因子模型以构造协方差矩阵 $\boldsymbol{\Sigma}_1$ 和 $\boldsymbol{\Sigma}_2$（请参见 Longford and Muthén, 1992; Poom and Lee, 1992; Longford, 1993, Linda et al., 1993; Muthén, 1994; Lee and Shi, 2001）。在群集层次的公共因子可被解释为群集层次的构念，它与单元层次的构念可能具有不同的因子结构。

要注意的是，传统上这些模型被称为双层次因子模型，因为单元被视为第一层次，群集被视为第二层次，且每个单元的响应向量被视为多元响应。如果将题器视为第一层次的单元，即如同我们的 GF 和 GRC 方程式那样，模型则变成了一个单元处在第二层次、群

集处在第三层次的三层次模型。为了简化起见，假设在第二层次和第三层次分别存在单独公共因子 $\eta_{1jk}^{(2)}$ 和 $\eta_{1k}^{(3)}$。第三层次的独特因子表示为 $\eta_{mk}^{(3)}, m=2,\cdots,I+1$。如图 4.1 所示为当 $I=3$ 时的模型图。

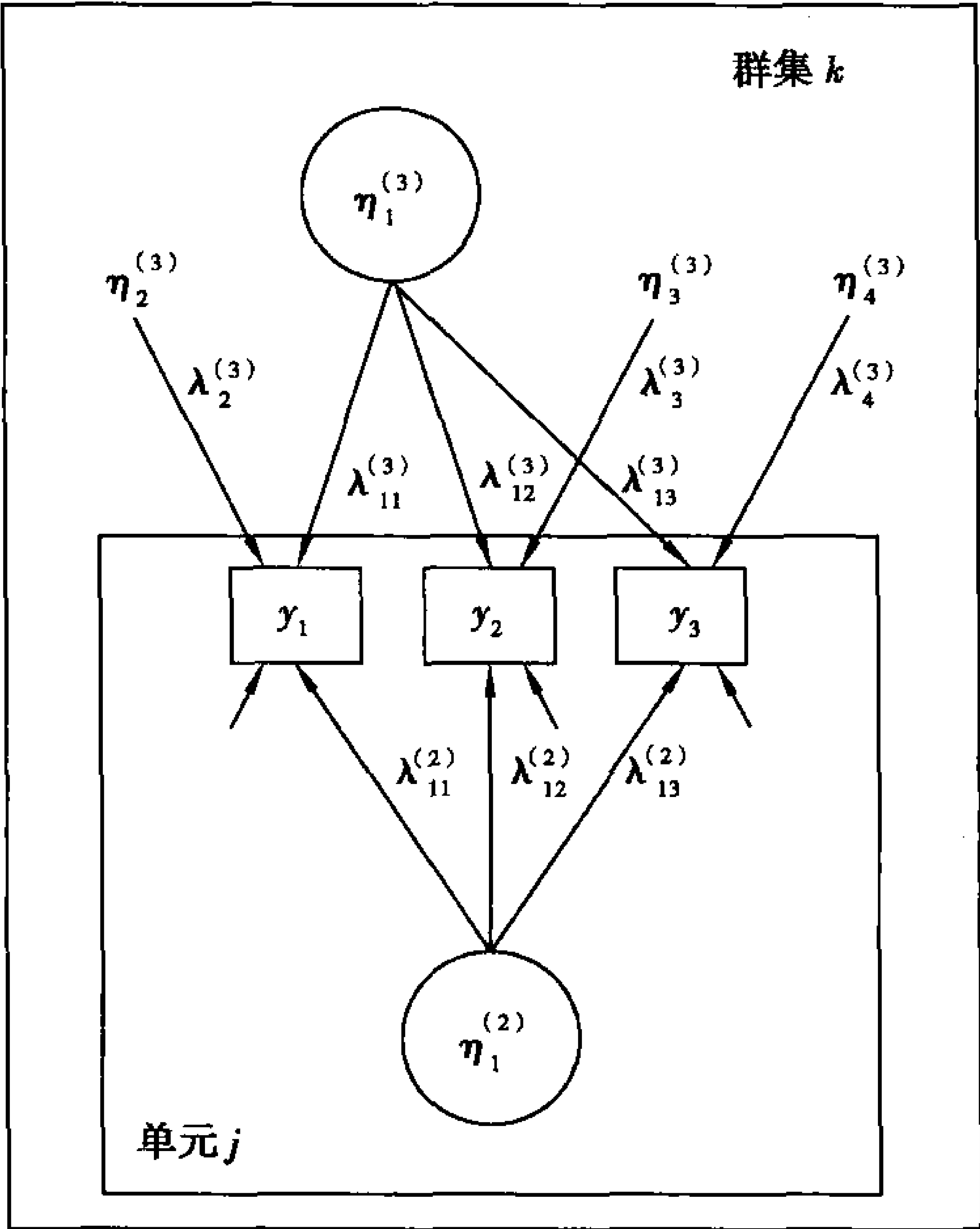



**图 4.1 “双层次”因子模型路径图**

图 4.1 中，内方框“单元 $j$”标记了题器以及第二层次的因子在单元层次上的变动情况。框中的模型是一个普通的单维因子模型，处在单元层次的因子以未标记的箭头表示，说明第一层次的变差性。外框中的潜变量在群集层次变动，用“群集 $k$”表示。这些潜变量表示一个被所有题器测量的公共因子，以及被每个题器单独测量的独特因子。与公共因子不同，独特因子没有用圆圈显示出来，因为它们仅仅表示残差项（没有因子负荷）。与传统的路径图相反，指向被观测响应 $y$ 的箭头，表示可能的非线性关系，如对数概率链接函数。指向被观测响应的未标示的短箭头并不一定表示额外的误差项，例如，它们也可能表示事件数的泊松变差性。

这个模型的 GF 方程在 4.2.4 节的展示 4.5A 中，其中，第三层次的单元 $k$ 含两个第二层次单元，$j=1,2$。要注意的是，图和展示中的因子负荷是根据展示 4.5B 中的 GRC 方程标记的。对于第 $i$ 个题器，每个公共因子都被乘以虚拟向量 $\mathbf{d}_i'$，第 $i$ 个元素等于 1，其他元素均为 0，这样，从 $\boldsymbol{\lambda}_1^{(3)}$ 和 $\boldsymbol{\lambda}_1^{(2)}$ 中可得到第 $i$ 个因子负荷。用 $d_{im}\lambda_{m+1}^{(3)}$ 乘以第 $m$ 个独特因子，当 $m=i$ 时，$d_{im}\lambda_{m+1}^{(3)}$ 等于 $\lambda_{i+1}^{(3)}$，否则等于 0。要注意的是，为了辨识这个模型，参数约束是必需的，如将因子负荷 $\lambda_{11}^{(2)}, \lambda_{11}^{(3)}, \lambda_2^{(3)}, \lambda_3^{(3)}, \lambda_4^{(3)}$ 设定为 1。在 5.2 节中，关于辨识有详细的论述。此模型的简化版本在 4.3.1 节中也有所介绍，另见图 4.3(a)。

### 4.2.5 例：GF 和 GRC 方程的探索性潜类模型

到目前为止，我们一直暗含的假设是潜变量是连续性的。尽管如此，仍能将同样的响应模型和离散潜变量结合起来以定义潜类模型。这里，双层次模型大体上已经足够了。定义 $\boldsymbol{\eta}_j^{(2)}$ 取值为离散值 $\mathbf{e}_c$，其概率为 $\pi_c$，并将均值约束为 0，即

$$\sum_c \pi_c \mathbf{e}_c = \mathbf{0}$$

对于二项响应的探索性潜类模型，题器为 $i$、单元为 $j$ 以及类别为 $c$ 的线性估计量可以表达为

$$\nu_{ijc} = \beta_i + e_{ic}, \qquad i = 1, \cdots, I$$

采用多元 GF 方程，公式变为

$$\boldsymbol{\nu}_{j(2)} = \boldsymbol{\beta} + \boldsymbol{\eta}_j^{(2)} \tag{4.10}$$

其中，$\boldsymbol{\eta}_j^{(2)} = \mathbf{e}_c, c = 1, \cdots, C$，式(4.10)是一个 $I$ 维潜变量，在这个例子中，$\mathbf{X}_{j(2)} = \mathbf{I}$，且 $\boldsymbol{\Lambda}_{j(2)} = \mathbf{I}$。

矩阵 $\mathbf{I}$ 是一个 $I$ 维单位矩阵。采用 GRC 方程式，线性估计量为

$$\nu_{ij} = \mathbf{d}_i'\boldsymbol{\beta} + \sum_{m=1}^{I} \eta_{jm}^{(2)} d_{mi} \tag{4.11}$$

其中，$\mathbf{d}_i'$是 $I$ 维单元矩阵的第 $i$ 行，且 $d_{mi}$是 $\mathbf{d}_i$ 的第 $m$ 个元素，即表示 $m = i$ 的虚拟变量。

对于题器 $i$ 含有 $S_i$ 个分类的多项响应或者比较响应的情形，那么单元 $j$，题器 $i$，响应类别 $s$，潜类 $c$ 的线性估计量为

$$\nu_{sijc} = \beta_i^s + e_{ic}^s$$

其中，对于 $i = 1, \cdots, I, c = 1, \cdots, C, \beta_i^1 = e_{ic}^1 = 0$。我们把不同响应类别和题器的线性估计量集中到一个单独向量 $\boldsymbol{\nu}_{j(2)}$ 中。那么，模型可重写为式(4.10)那样，但是 $\boldsymbol{\eta}_j^{(2)}$ 的维度为 $R = \sum_i S_i - I$，并且单位矩阵被替换为 $(R+I) \times R$ 维的结构矩阵，后者相当于从 $(R+I) \times (R+I)$ 维的单位矩阵移除了各题器所对应的第 1 响应类别的 $I$ 列。在式(4.1.1)中的 GRC 方程中，$\mathbf{d}_i'$被这个结构矩阵中的第 $i$ 行所代替。在 9.3 节，我们将使用一个探索性潜类模型来分析二项诊断测试，在 13.5 节分析政治价值排序。

---

展示 4.5 双层次因子模型的随机项部分

---

含两个二层次单元 $j = 1, 2$ 的三层次单元 $k$：

A. GF 方程：

$$\boldsymbol{\Lambda}_{k(3)}^{(2)}\boldsymbol{\eta}_{k(3)}^{(2)} + \boldsymbol{\Lambda}_{k(3)}^{(3)}\boldsymbol{\eta}_{k(3)}^{(3)} = \underbrace{\left[\begin{array}{cc} \lambda_{11}^{(2)} & 0 \\ \vdots & \vdots \\ \lambda_{1I}^{(2)} & 0 \\ \hline 0 & \lambda_{11}^{(2)} \\ \vdots & \vdots \\ 0 & \lambda_{1I}^{(2)} \end{array}\right]}_{\boldsymbol{\Lambda}_{k(3)}^{(2)}} \underbrace{\left[\begin{array}{c} \eta_{11k}^{(2)} \\ \hline \eta_{12k}^{(2)} \end{array}\right]}_{\boldsymbol{\eta}_{k(3)}^{(2)}} + \underbrace{\left[\begin{array}{cccc} \lambda_{11}^{(3)} & \lambda_{2}^{(3)} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ \lambda_{1I}^{(3)} & 0 & \cdots & \lambda_{I+1}^{(3)} \\ \hline \lambda_{11}^{(3)} & \lambda_{2}^{(3)} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ \lambda_{1I}^{(3)} & 0 & \cdots & \lambda_{I+1}^{(3)} \end{array}\right]}_{\boldsymbol{\Lambda}_{k(3)}^{(3)}} \underbrace{\left[\begin{array}{c} \eta_{1k}^{(3)} \\ \eta_{2k}^{(3)} \\ \vdots \\ \eta_{I+1,k}^{(3)} \end{array}\right]}_{\boldsymbol{\eta}_{k(3)}^{(3)}}$$

$$\boldsymbol{\Lambda}_{k(3)}\boldsymbol{\eta}_{k(3)} = \underbrace{\left[\begin{array}{cccccc} \lambda_{11}^{(2)} & 0 & \lambda_{11}^{(3)} & \lambda_{2}^{(3)} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ \lambda_{1I}^{(2)} & 0 & \lambda_{1I}^{(3)} & 0 & \cdots & \lambda_{I+1}^{(3)} \\ \hline 0 & \lambda_{11}^{(2)} & \lambda_{11}^{(3)} & \lambda_{2}^{(3)} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & \lambda_{1I}^{(2)} & \lambda_{1I}^{(3)} & 0 & \cdots & \lambda_{I+1}^{(3)} \end{array}\right]}_{\boldsymbol{\Lambda}_{k(3)}} \underbrace{\left[\begin{array}{c} \eta_{11k}^{(2)} \\ \eta_{12k}^{(2)} \\ \eta_{1k}^{(3)} \\ \eta_{2k}^{(3)} \\ \vdots \\ \eta_{I+1,k}^{(3)} \end{array}\right]}_{\boldsymbol{\eta}_{k(3)}}$$

B. GRC 方程：

$$\eta_{1jk}^{(2)}\mathbf{z}_{1i}^{(2)\prime}\boldsymbol{\lambda}_1^{(2)}+\sum_{m=1}^{l+1}\eta_{mk}^{(3)}\mathbf{z}_{mi}^{(3)\prime}\boldsymbol{\lambda}_m^{(3)}=\eta_{1jk}^{(2)}\underbrace{\mathbf{d}_i'}_{\mathbf{z}_{1i}^{(2)\prime}}\underbrace{\begin{bmatrix}\lambda_{11}^{(2)}\\ \vdots \\ \lambda_{1l}^{(2)}\end{bmatrix}}_{\boldsymbol{\lambda}_1^{(2)}}+\eta_{1k}^{(3)}\underbrace{\mathbf{d}_i'}_{\mathbf{z}_{1i}^{(3)\prime}}\underbrace{\begin{bmatrix}\lambda_{11}^{(3)}\\ \vdots \\ \lambda_{1l}^{(3)}\end{bmatrix}}_{\boldsymbol{\lambda}_1^{(3)}}+\eta_{2k}^{(3)}\underbrace{d_{i1}}_{z_{2i}^{(3)}}\lambda_2^{(3)}+\cdots+\eta_{l+1,k}^{(3)}\underbrace{d_{i,l}}_{z_{l+1,i}^{(3)}}\lambda_{l+1}^{(3)}$$

$$=\eta_{1jk}^{(2)}\lambda_{1i}^{(2)}+\eta_{2k}^{(3)}\lambda_{1i}^{(3)}+\eta_{i+1,k}^{(3)}\lambda_{i+1}^{(3)}$$

其中，$d_{ir}=\begin{cases}1 & 当\ r=i\\ 0 & 其他\end{cases}$ 且，$\mathbf{d}_i'=(d_{i1},d_{i2},\cdots,d_{il})$。

---

## 4.2.6 释放条件独立假设

模型框架的一个基本假定是在给定潜变量和协变量的条件下，响应是条件独立性的，这也被称为“局部独立”假设。这个假设可能会显得过于严格，但是通过引入更多潜变量，将得到更为复杂的结构。例如，在一个公共因子模型中，可通过将响应负荷在另外一个潜变量上并设其负荷为1，在两个响应之间建立相关关系，并使之条件独立于公共因子。同样的，曲等人(Qu et al.,1996)也在一个潜类模型中释放了条件独立假设，他们引入一个所有题器的公共因子，并且各题器的因子负荷在每个类别上都不同。

设 $\epsilon_{1j}$是正态概率回归中潜响应的残差；$\epsilon_{2j}$是线性回归中观测响应的残差，正如著名的赫克曼选择模型(Heckman selection model)(请参见 Heckman,1979)，为了在残差 $\epsilon_{1j}$和 $\epsilon_{2j}$之间引入非独立关系，可设定

$$\epsilon_{ij} = \eta_j^{(2)}\lambda_i + e_{ij},\quad \lambda_1 = 1$$

其中，第一层次残差 $e_{1j}$和 $e_{2j}$服从均值为0，方差分别为 $\theta_{11}$和 $\theta_{22}$的独立正态分布。由于仅有线性回归模型的残方差

$$\text{Var}(\epsilon_{2j}) = \lambda_2^2\psi + \theta_{22} \tag{4.12}$$

以及这两个模型的总体残差的相关系数

$$\text{Cor}(\epsilon_{1j},\epsilon_{2j}) = \frac{\lambda_2\psi}{\sqrt{(\psi+\theta_{11})(\lambda_2^2\psi+\theta_{22})}}$$

可以被辨识(见5.2节模型辨识的一般处理)，因此，需要在4个参数($\psi,\lambda_2,\theta_{11}$and $\theta_{22}$)之外再增加两个约束。这里不能将 $\lambda_2$ 设为常量，因为这就确定了相关的方向。因而，很明确的一个选择是，设 $\theta_{11}=1$(正如正态概率回归中常用的方法)以及 $\psi=1$。然而如此一来，两个模型的总体残的相关系数则变为

$$\text{Cor}(\epsilon_{1j},\epsilon_{2j}) = \frac{\lambda_2}{\sqrt{(2)(\lambda_2^2+\theta_{22})}} \leqslant \frac{\lambda_2}{\sqrt{2\lambda_2^2}} = \frac{1}{\sqrt{2}}$$

其中，如果线性回归模型中的 $\theta_{22}=0$，则会存在上限。为了避免相关的上限(小于1)，可以设约束参数为

$$\psi = 1,\quad \theta_{11} = \theta_{22}$$

对于类别响应变量，可以释放条件独立性，而无须在模型中引入更多潜变量。考察一下两个二项响应 $y_1$ 和 $y_2$，将4个可能的响应模式(0,0)，(1,0)，(0,1)，(1,1)作为一个单独的多项响应，建模为

$$\Pr(y_1,y_2) = \frac{\exp(\beta_1 y_1+\beta_2 y_2+\beta_{12}y_1 y_2)}{\sum_{z_1=0,1}\sum_{z_2=0,1}\exp(\beta_1 z_1+\beta_2 z_2+\beta_{12}z_1 z_2)}$$

除了由潜变量所引入的额外相关以外，其他的额外相关都能在 $\beta_{12} \neq 0$ 的条件下有解。这种在潜类模型中引入“局部”依赖的方法由哈珀(Harper,1972)和哈格纳斯(Hagenaars,1988)以及其他一些学者所提出，但一般来说，也适用于含类别响应的潜变量模型中。例如，在题器响应建模中，一组题器中的局部依赖，可通过将题器汇总合并到一个称之为“测试组”(Wainer and Kiely,1987)或“题器束(item bundle)”(Wilson and Adams,1995)的单一响应中来处理。

## 4.3 潜变量的结构模型

### 4.3.1 连续潜变量

为了定义结构模型，首先定义潜变量向量 $\eta_{j\cdots z} = (\boldsymbol{\eta}_{jk\cdots z}^{(2)\prime}, \boldsymbol{\eta}_{k\cdots z}^{(3)\prime}, \cdots, \boldsymbol{\eta}_{z}^{(L)\prime})'$，它包含了**第 $j$ 个第二层次单元的所有潜变量**，其中，$k,\cdots,z$ 是第三层次到第 $L$ 层次的单元索引；向量也可以用 $\boldsymbol{\eta}_{jk\cdots z(2)}$ 或 $\boldsymbol{\eta}_j$ 表示。对于一个三层次模型，当 $j$ 取值为 2 时，4.2.3 节中展示 4.4 中的左上图即为 $\boldsymbol{\eta}_{jk}$。潜变量的结构模型具有如下形式(省略了较高层次的下标)：

$$\boldsymbol{\eta}_j = \mathbf{B}\boldsymbol{\eta}_j + \boldsymbol{\Gamma}\mathbf{w}_j + \boldsymbol{\zeta}_j \tag{4.13}$$

其中，$\mathbf{B}$ 是一个 $M \times M$ 的回归参数矩阵，$M = \sum_l M_l$，$\mathbf{w}_j$ 是包含 $R$ 个协变量的向量，$\boldsymbol{\Gamma}$ 是一个 $M \times R$ 的回归参数矩阵，$\boldsymbol{\zeta}_j$ 是具有 $M$ 个误差或干扰项的向量。这个模型本质上是从传统的单层次结构模型(请参见 Muthén,1984)到多层次的一般扩展。但其关键不同在于，在我们的模型框架中，潜变量和观测变量在不同的层次上可以发生变动。$\boldsymbol{\zeta}_j$ 的每个元素，与对应的 $\boldsymbol{\eta}_j$ 的元素一样，在相同的层次上变动。

一个较高层次的潜变量对一个较低层次的潜变量或者观测变量进行回归是没有意义的，因为这可能会迫使高层次变量在低层次上变动。如果 $\mathbf{B}$ 的矩阵块对应于每个层次上的潜变量向量 $\boldsymbol{\eta}^{(l)}$，那么矩阵 $\mathbf{B}$ 就是一个上分块对角矩阵。同样的，如果协变量向量 $\mathbf{w}_j$ 表示为 $\mathbf{w}_j = (\mathbf{w}_{jk\cdots z}^{(2)\prime}, \mathbf{w}_{k\cdots z}^{(3)\prime}, \cdots, \mathbf{w}_{z}^{(L)\prime})$，那么矩阵 $\boldsymbol{\Gamma}$ 也是一个上分块对角矩阵：

$$\begin{bmatrix} \boldsymbol{\eta}_{jk\cdots z}^{(2)} \\ \boldsymbol{\eta}_{k\cdots z}^{(3)} \\ \vdots \\ \boldsymbol{\eta}_{z}^{(L)} \end{bmatrix} = \begin{bmatrix} \mathbf{B}^{(22)} & \mathbf{B}^{(23)} & \cdots & \mathbf{B}^{(2L)} \\ \mathbf{0} & \mathbf{B}^{(33)} & \cdots & \mathbf{B}^{(3L)} \\ \vdots & \vdots & & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{B}^{(LL)} \end{bmatrix} \begin{bmatrix} \boldsymbol{\eta}_{jk\cdots z}^{(2)} \\ \boldsymbol{\eta}_{k\cdots z}^{(3)} \\ \vdots \\ \boldsymbol{\eta}_{z}^{(L)} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\Gamma}^{(22)} & \boldsymbol{\Gamma}^{(23)} & \cdots & \boldsymbol{\Gamma}^{(2L)} \\ \mathbf{0} & \boldsymbol{\Gamma}^{(33)} & \cdots & \boldsymbol{\Gamma}^{(3L)} \\ \vdots & \vdots & & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\Gamma}^{(LL)} \end{bmatrix} \begin{bmatrix} \mathbf{w}_{jk\cdots z}^{(2)} \\ \mathbf{w}_{k\cdots z}^{(3)} \\ \vdots \\ \mathbf{w}_{z}^{(L)} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\zeta}_{jk\cdots z}^{(2)} \\ \boldsymbol{\zeta}_{k\cdots z}^{(3)} \\ \vdots \\ \boldsymbol{\zeta}_{z}^{(L)} \end{bmatrix} \tag{4.14}$$

分块 $\mathbf{B}^{(ab)}$ 包含了 $\boldsymbol{\eta}^{(a)}$ 对 $\boldsymbol{\eta}^{(b)}$ 回归的回归参数，$\boldsymbol{\Gamma}^{(ab)}$ 也是如此。注意，尽管如此，与 $\boldsymbol{\eta}_j$ 不同，$\mathbf{w}_j$ 不需要包含各层次的子向量。例如，在 $L$ 层次上可能有一个单个协变量($R=1$，$\mathbf{w}_j = w_z^{(L)}$)。本书后续将省略协变量 $w$ 的上标。

如果在某个层次的潜变量之间的回归是**递归**的，模型就较容易估计和理解。在上面的例子中，可以按照这样的原则改变 $\boldsymbol{\eta}^{(l)}$ 中元素的序列，这个原则是，分块 $\mathbf{B}^{(aa)}$ 必须严格作为上分块对角。这样，$\eta_{M_l}^{(l)}$ 的表达式可以替换成 $\eta_1^{(l)}$ 的表达式，直到 $\eta_{M_l-1}^{(l)}$，$\eta_{M_l-1}^{(l)}$ 的表达式可以替换成 $\eta_1^{(l)}$ 表达式，直到 $\eta_{M_l-2}^{(l)}$，以此类推，直到等式右边所有的 $\eta_m^{(l)}$ 都被消除。将最后的表达式替换为线性估计量，然后产生出潜变量的简化式(见 4.6 节)，在这个简化式中，剩余在右边的唯一的潜变量是干扰项 $\boldsymbol{\zeta}$。将变量关系限定为递归的做法隐含着这样的事实，即它不能与回归到另一个潜变量的特殊潜变量有同步的效应，反之亦然。尽

管如此,这样的模型很少在实践中使用,可能是由于概念上的复杂和模型辨识上的局限。

**例子**

有这样一个结构方程,它包括两个第二层次潜变量和一个第三层次潜变量,假设如下:

$$\begin{bmatrix}\eta_{1jk}^{(2)}\\ \eta_{2jk}^{(2)}\\ \eta_{1k}^{(3)}\end{bmatrix}=\begin{bmatrix}0 & b_{12} & b_{13}\\ 0 & 0 & 0\\ 0 & 0 & 0\end{bmatrix}\begin{bmatrix}\eta_{1jk}^{(2)}\\ \eta_{2jk}^{(2)}\\ \eta_{1k}^{(3)}\end{bmatrix}+\begin{bmatrix}\gamma_{11} & 0\\ 0 & \gamma_{22}\\ 0 & 0\end{bmatrix}\begin{bmatrix}w_{jk}\\ w_{k}\end{bmatrix}+\begin{bmatrix}\zeta_{1jk}^{(2)}\\ \zeta_{2jk}^{(2)}\\ \zeta_{1k}^{(3)}\end{bmatrix}$$

其路径图如图4.2所示。其中,$\eta_{1jk}^{(2)}$对相同层次的潜变量$\eta_{2jk}^{(2)}$,较高层次的潜变量的$\eta_{1jk}^{(3)}$,以及在第二层次变动的观测协变量$w_{jk}$回归。逆转$b_{13}$到$b_{31}$的路径没有意义,因为$\eta_{1k}^{(3)}$将被迫在第二层次上变动,而增加一条从$\eta_{1jk}^{(2)}$到$\eta_{2jk}^{(2)}$的路径$b_{21}$,将使第二层次上的关系变成非递归关系。

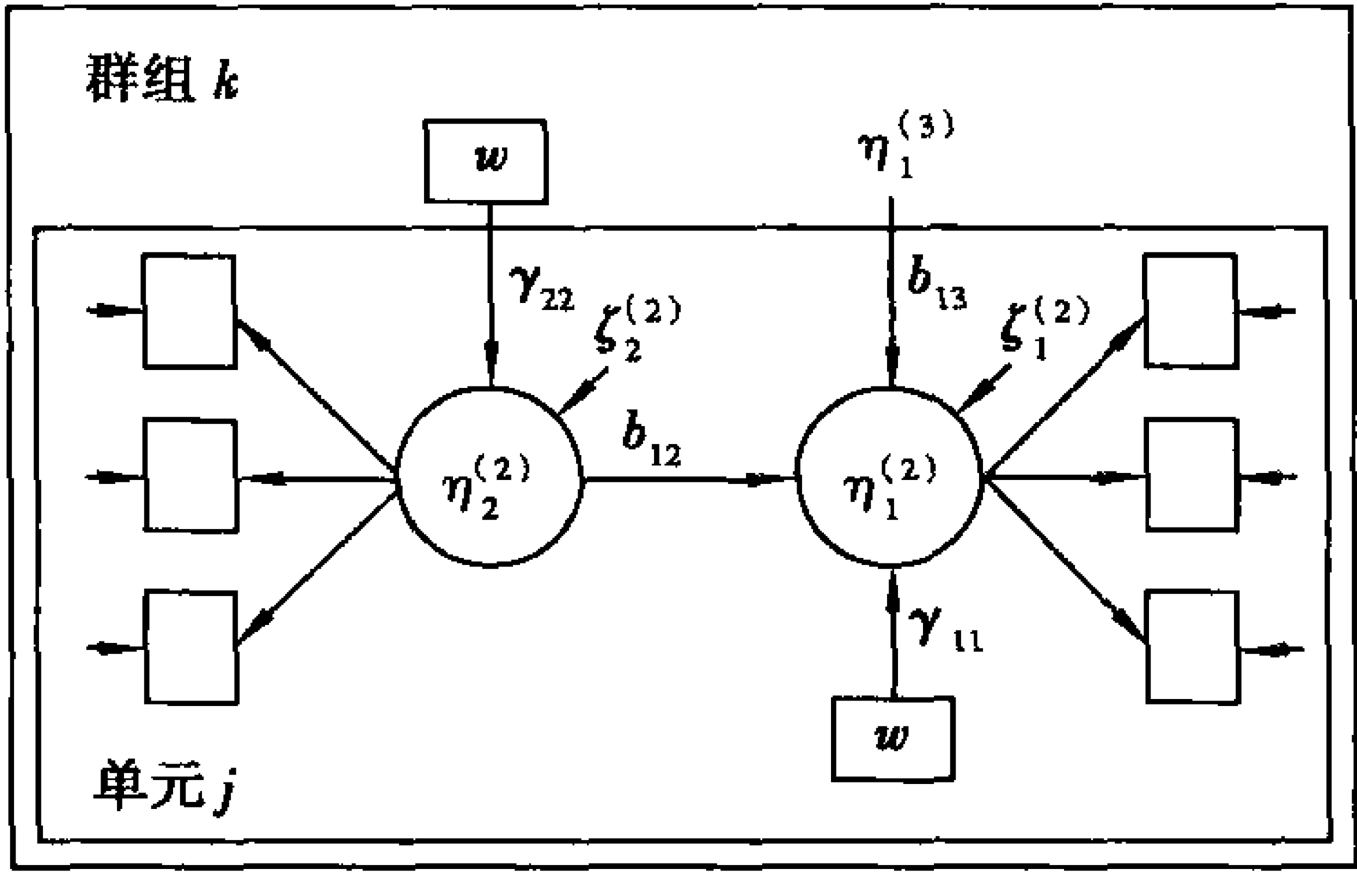



**图4.2 多层次结构方程模型图例**

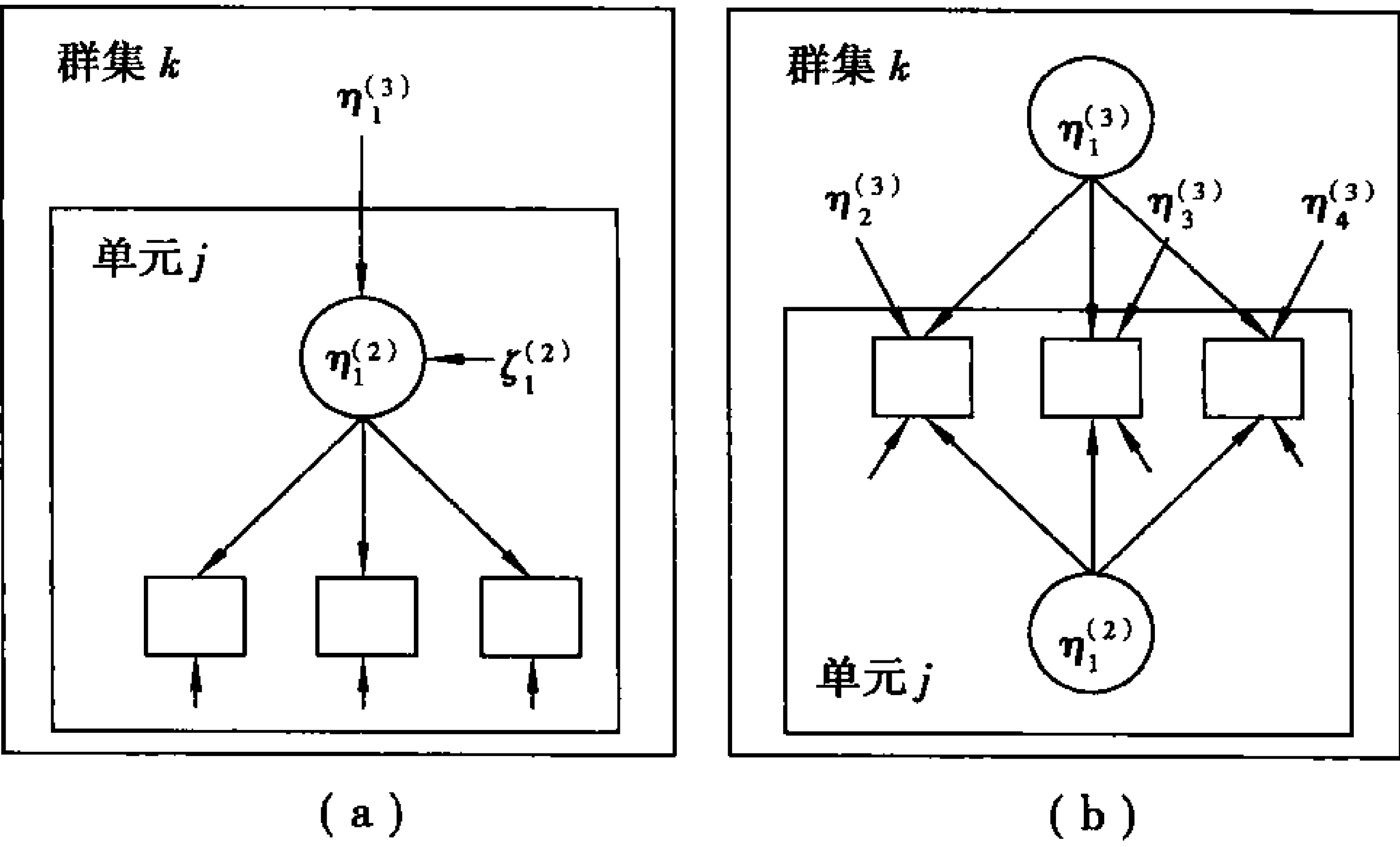
现在将考察4.2.4节提到的双层次因子模型的另一个方案,图4.3(b)也有所展示。如图4.3(a)所示,维持双层次模型(内框部分)原有的特性,通过在第二层次因子与第三层次因子间建立回归关系的方法,可以使第二层次因子在第三层次变动。这样的模型即为**方差成分因子模型**(variance components factor model)(Rabe-Hesketh et al.,2004a)。可以证明,这个模型比图4.3(b)中的结构化程度较弱的方案更容易去解释。对于公共因子,可以通过考察它与单元层次观测题器的关系进行定义,可知,在群集层次上,它仅仅有导致变差性的成分。这个模型与MIMIC类似,其中最大的差异是,前者的公共因子对在较高层次中变动的潜变量回归,而非与观测变量。含有在较高层次的独特因子与在MIMIC模型中含有直接效应是相类似的。方差成分因子模型的一个优势是,它比一般双层次因子模型更容易在结构等式模型内进行合并。事实上,这个模型部分的形成为图4.2中例子的模型,在前面的例中,$\eta_{1jk}^{(2)}$是一个第二层次的公共因子,$\eta_{1k}^{(3)}$是第三层次的方差成分。方差成分因子模型可用来分析9.8节中对流产的态度问题。

非线性潜变量结构模型也已经被提出来(请参见,Busemeyer and Jones,1983;Kenny and Judd,1984)。阿明戈尔和穆森(Arminger and Muthén,1998)论述了连续响应的LISREL模型的非线性版(见3.5节)。依据潜在探索性变量$\boldsymbol{\zeta}_j$,潜响应变量$\boldsymbol{\eta}_j$结构模型可以表达为

$$\boldsymbol{\eta}_j=\mathbf{B}\boldsymbol{\eta}_j+\boldsymbol{\Gamma}\boldsymbol{\alpha}_j+\boldsymbol{\zeta}_j,\quad \boldsymbol{\alpha}_j=\mathbf{g}(\boldsymbol{\xi}_j)$$

其中,$\mathbf{g}(\boldsymbol{\xi}_j)$是一个已知的确定的向量,且假设$\boldsymbol{\xi}_j$和$\boldsymbol{\zeta}_j$服从正态分布。特殊的例子还包括多项回归模型,其中,

$$\boldsymbol{\alpha}_j=[\xi_j,\xi_j^2,\cdots,\xi_j^p]$$




(a)　　　　(b)

**图**4.3

(a)方差成分因子模型　(b)一般双层次因子模型

以及(一阶)交互作用模型,其中

$$\boldsymbol{\alpha}_j = \left[\xi_{1j},\xi_{2j},\cdots,\xi_{qj},\xi_{1j}\xi_{2j},\xi_{1j}\xi_{3j},\cdots,\xi_{q-1,j}\xi_{qj}\right]$$

设 $\boldsymbol{\eta}_j=\boldsymbol{\alpha}_j$,则产生一个非线性因子模型。我们建议参考杰拉斯科格(Jöreskog,1998),以了解非线性结构方程建模研究的总体情况,另可参考 Schumaker and Marcoulides(1998)。

## 4.3.2　离散潜变量

对于离散潜变量来说,结构模型是(先验)概率模型,即其各个单元归属于相应的潜类别的概率。对于一个第二层次的单元 $j$,设其属于类别 $c$ 的概率用 $\pi_{jc}$表示,则

$$\pi_{jc} \equiv \Pr(\boldsymbol{\eta}_j^{(2)} = \mathbf{e}_c)$$

这个概率可通过一个多项对数概率模型而依赖于协变量 $\mathbf{w}_j$,

$$\pi_{jc} = \frac{\exp(\mathbf{w}_j'\boldsymbol{\varrho}^c)}{\sum_d \exp(\mathbf{w}_j'\boldsymbol{\varrho}^d)} \tag{4.15}$$

其中,$\boldsymbol{\varrho}^c$ 是回归参数,为模型辨识起见,设 $\boldsymbol{\varrho}^1=\mathbf{0}$。这样的“伴随变量(concomitant variable)”潜类模型在戴顿和麦克里迪(Dayton and MacReady,1988)以及福尔曼(Formann,1992)的例子中也出现过。即使类别隶属并不依赖于协变量,多项对数概率参数化也有很大用处,因为它使得潜类别概率总和为 1。

使用下面的约束条件往往对于原点有所帮助。用 $\pi_c$表示当协变量 $\mathbf{w}_j$ 等于 0 时(除了常数)的概率。那么对于 $c=1,\cdots,C-1$,$\mathbf{e}_c$ 是自由参数,并且通过设定均值原点为 0 可确定 $\mathbf{e}_c$,即

$$\sum_{c=1}^{C}\pi_c\mathbf{e}_c = \mathbf{0}$$

这种参数化的优点在于,均值结构可以设定在模型 $\mathbf{x}_{ij}'\beta$ 中的固定部分,如同在连续潜变量模型中那样。

如果潜类别按照某个维度排序,如在离散单因子模型中,则通过约束模型(4.15)中的参数或者使用累积模型(见 2.3.4 节),定序模型也可被设定为潜类概率。

如果有数个离散潜变量,即可使用模型(4.15),其中,$c$ 标记了潜变量类别的合并;也可使这个模型参数化为一个含有主效应以及潜变量交互项的对数线性模型。哈格纳斯(Hagenaars,1993)和沃芒(Vermunt,1997)考察了在同一个层次上,离散潜变量与其他离散潜变量的回归关系。沃芒(Vermunt,2003)扩展了结构模型,使得模型(4.15)中的线性

预测方程包括较高层次的连续或者离散潜变量，可参见 4.4.4 节。

## 4.4 干扰项分布

为了完成模型设定，在结构模型中，必须设定干扰项 $\boldsymbol{\zeta}$ 的分布。如果没有结构模型，那么潜变量就等于干扰项，即 $\boldsymbol{\eta}=\boldsymbol{\zeta}$。

干扰项的依赖结构通过层次的个数 $L$ 以及在每个层次中干扰项的个数 $M_l$ 来设定。一个特殊层次可能与层级数据中的某个群集层次相重合。尽管如此，在模型层次和数据等级的层次之间，并不经常有直接的对应关系。例如，在因子模型中，模型题器被视为在第一层次的单元，而受试者则被视为第二层次的单元。

"某层次单元""某层次干扰项"是这样定义的：

- 一个处于第一层次的单元是观测的基本单元。
- 处于 $l$ 层次的单元 $k$($l$ 大于 1)是第一层次单元的群集。
- 处于 $l(l>1)$层次的群集 $k$ 中的第一层次单元，包含于 $n_k^{(l-1)}$ 层次的单元组成的子集 $n_k^{l-1}$，后者表示 $l-1$ 层次的单元。
- 处于 $l$ 层次的干扰项 $\zeta^{(l)}$ 在 $l$ 层次的单元之间变动，但不在单元之内。
- 给定处于 $l+1$ 层次或更高层次的干扰项以及任一解释变量，$l$ 层次的单元条件独立的。

基本的假设是，处于同一层次上的干扰项可能是相互依赖的，而不同层次上的干扰项则是独立的。在下面的小节中，将会介绍分布 $\boldsymbol{\zeta}$ 的不同设定方式。

### 4.4.1 连续性分布

如果是连续性干扰项，主要的分布假设是均值为 0，且第 $l$ 层次的协方差矩阵为 $\boldsymbol{\Psi}^{(l)}$ 的多元正态分布。这种分布的优点是，均值、方差、协方差都被参数化，可以自由设定。很重要的一点是，在这种情况下，如果响应变量不为条件正态分布时，似然值则无法表示为闭式。尽管如此，如在 6.2 中论述过的，如果是含有"组间"协变量的简单随机截距模型，潜变量和响应分布合并，闭式表达式才能存在。维德尔和鎌仓(Wedel and Kamakura, 2001)论述了含独立因子的因子模型，它包含来自于指数族的任意分布，尽管这些模型通常缺少闭式似然。

幸运的是，在许多情况下，即使干扰项偏离正态分布，其推导依然出乎意料的稳健(请参见 Bartholomew, 1988, 1994; Seong, 1990; Kirisci and Hsu, 2001; Wedel and Kamakura, 2001)。同时，很多人尝试去寻找更加具有"稳健性"的干扰项分布。皮涅罗等(Piheiro et al., 2001)考察了多元 $t$ 分布，发现在处理异常值方面比多元正态分布更为稳健。

为了避免对干扰项的分布做出硬性的假设，可以用灵活度更高的参数分布，如(多元)正态分布的有限混合(请参见 Uebersax, 1993; Uebersax and Grove, 1993; Magder and Zeger, 1996; Verbeke and Lesaffre, 1996; Allenby et al., 1998; Carroll et al., 1999; Lenk and DeSarbo, 2000; Richardson et al., 2002)。在某些情形下，有限混合的成分被解释成子群体，如在曲等(Qu et al., 1996)中的"有疾病或无疾病的人"，具体见 4.4.4 节。按照贝叶斯定理设置，通过"狄利克雷进程(Dirichlet process)"的半参数混合进行干扰项建模，最近引起了较大的关注(请参见 Müller and Roeder, 1997; Chib and Hamilton, 2002)。

另一种方法是采用加兰特和尼驰卡(Gallant and Nychka, 1987)、大卫迪安和加兰特

(Davidian and Gallant,1992)提出的"截断埃尔米特级数(truncated Hermite series)"法。

### 4.4.2 非参数分布

如果不建立关于干扰项的分布假设,可以采用"非参数最大似然估计法"(NPMLE)(Laird,1978)。基本上,在随机截距模型的前后关系中,希玛(Simar,1976)和莱尔德(Laird,1978)以及林赛(Lindsay,1983)已经指出,未设定分布(可能是连续性的分布)的非参数最大似然估值(NPMLE)是概率 $\pi_c$ 非零的离散分布,这个概率在原点 $\mathbf{e}_c$ 的有限组中,$c=1,\cdots,C$,如图 4.4 所示的上方框格(请参见 Lindsay et al.,1991;Aitkin,1996,1999a;Rabe-Hesketh,2003a),据此,此模型通常被视为非参数混合模型。从图 4.4 下方框中可知,干扰项的累积分布函数是一个阶梯函数(step function)。

对于一个含 $M$ 个元素的多元干扰项,质量位于 $M$ 维上的点 $\mathbf{e}_c$。(请参见 Davies and Pickles,1987;Aitkin,1999a)。例如,在 9.5 节中的例子 $M=2$,二维"质量点"分布以两种方式显示在图 9.7 中。沃芒(Vermunt,2004)介绍了一个三层次模型的非参数最大似然估计值。

质量的数目已知,使用最大似然估计可对其原点、概率及其他参数进行估计。质量的数目不断增加,直到最大化似然出现。或者,也可对一个质量数目很大的模型进行估计,这样使多余的质量或与其他的质量合并(有同一个原点),或者概率为 0。在 6.5 节,讨论了一种以方向导数为基础,来判定所给定的 $C$ 是否对应于 NPMLE 的方法。林赛(Lindsay,1995)和波宁(Böhning,2000)曾经对最大似然理论进行过总结回顾。埃南(Heinen,1996)认为 NPMLE 是"完全半参数的",并谈到了一种较为简单的方法,质量得以估计,但设定原点为先验的"半参数"。

NPMLE 一个较大的优点在于,它能适时地排除干扰分布的效应。真实的分布可能是连续性的(正态的或者非正态的),离散的或者连续但某些成分是离散的。依靠 NPMLE 可以将注意力放在其他模型分量的设定上,而不必担心干扰分布的性质。在 9.5,11.2,11.3.3,14.2 节中都使用了 NPMLE 作为例子。

### 4.4.3 离散分布

如果潜变量是离散的,可用多元对数概率模型建模,如在 4.3.2 节的论述。

### 4.4.4 混合连续分布和离散分布

同时含有连续和离散潜变量的模型可以采用以下不同的公式:

第一种在响应模型中,包括潜变量的连续和离散两种类型。在一个排序模型中,博肯哈特(Böckenholt,2001b)涵括了离散的特定选项的随机截距,以及连续性公因子和随机系数。同样,麦克拉奇等(McCulloch et al.,2002)设定了一个"潜类混合模型",用离散和连续随机系数进行连续纵贯响应和生存分析的联合建模(14.6 节有类似的应用)。潜类别被解释为群体的子群体,这些子群体在(对数)前列腺抗原的均值轨迹和前列腺癌的发作时间上都会有差别。同一个潜类别中,不同人之间的变差性能够通过(连续的)随机效应进行调适。博肯哈特(Böckenholt,2001b)和麦克拉奇(McCulloch,2002)两个人都将离散和连续潜变量视为互相独立的。注意,离散和连续(均值为 0,正态分布)的潜变量的总数,仅仅是正态密度的有限混合,两个潜变量方差相同,离散变量表示成分均值。

第二种类型的模型在响应模型中仅有离散潜变量,在结构模型中含连续潜变量。这种模型由沃芒(Vermunt,2003)在模型多级设置中提出。在题器-层次的模型是一个条件

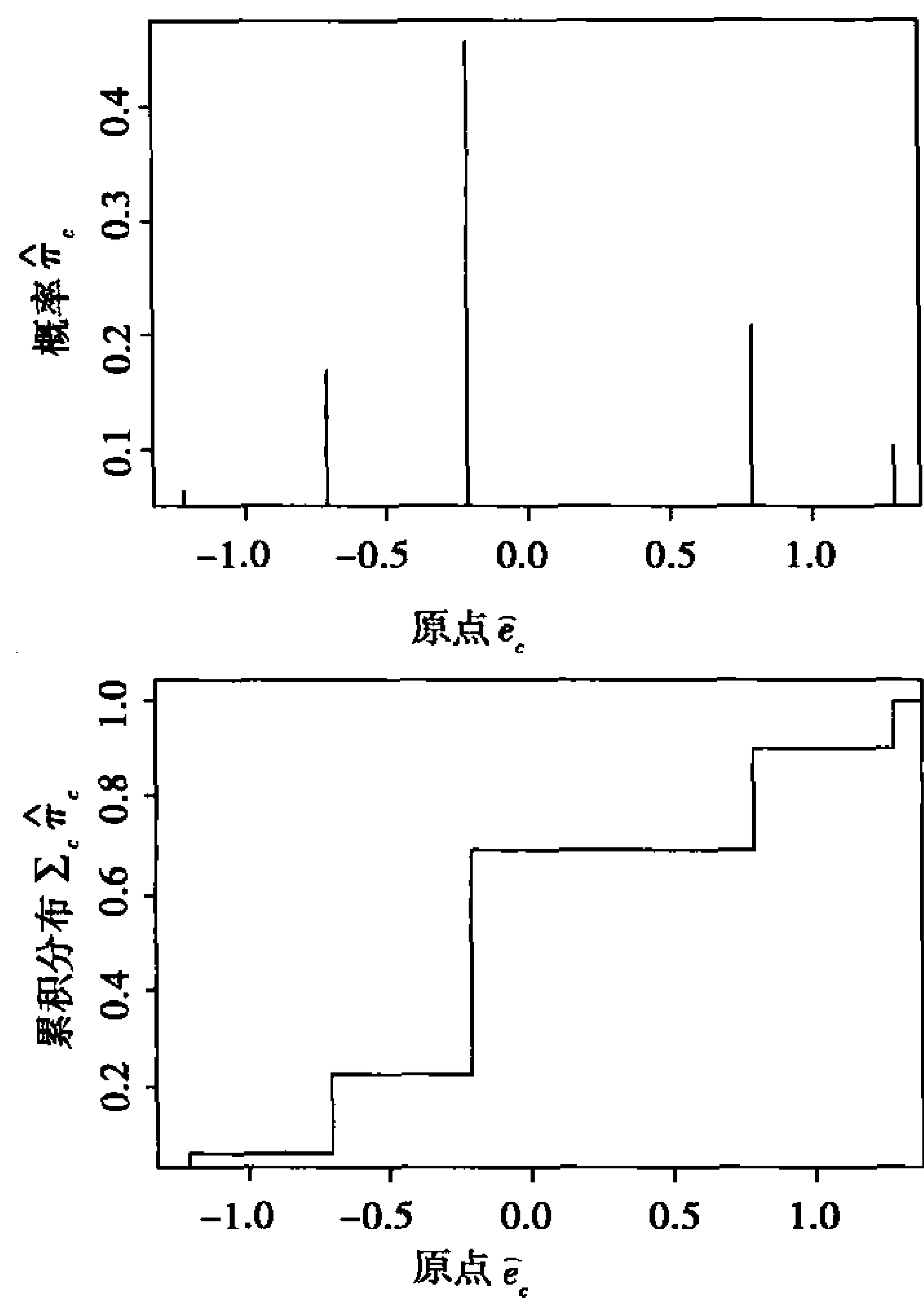



**图 4.4　离散分布和累积分布**

响应模型,其单元为 $j$,题器为 $i$,群集为 $k$,类别隶属为 $c$。如果是含有分类 $s=1,\cdots,S$ 的无序类别响应,则模型可以表达为

$$\Pr(y_{ijk}=s\mid \boldsymbol{\eta}_{jk}^{(2)}=\mathbf{e}_c)=\frac{\exp(\nu_{ijkc}^{s})}{\sum_{t=1}^{S}\exp(\nu_{ijkc}^{t})} \tag{4.16}$$

其中,$\nu_{ijkc}^{s}$ 是分类 $s$ 的线性估计量。单元-层次模型是一个类别隶属的多项对数概率模型,即

$$\Pr(\boldsymbol{\eta}_{jk}^{(2)}=\mathbf{e}_c)=\frac{\exp(\alpha_{jk}^{c})}{\sum_{b}\exp(\alpha_{jk}^{b})}$$

其中,结构模型的线性估计量 $\alpha_{jk}^{c}$ 包含一个群集层次随机截距,即

$$\alpha_{jk}^{c}=\mathbf{w}_{jk}'\boldsymbol{\varrho}^{c}+\boldsymbol{\eta}_{k}^{(3)}$$

其中,$\mathbf{w}_{jk}$ 是特定单元和特定群集的协变量,含固定的类别系数 $\boldsymbol{\varrho}^{c}$,将群集层次的随机截距设为正态分布。沃芒认为假设条件响应概率不依存于群集(如式(4.16),通过舍弃下标 $k$)通常很有用处。他同样指出,群集层次随机截距可以设为离散。

第三种在响应模型中仅仅含有连续潜变量,而离散潜变量则在结构模型中出现。在结构模型中,连续变量对离散潜变量回归,这种模型常常要比传统的结构模型(仅仅包括连续潜变量)复杂,如残差协方差矩阵可能取决于离散潜变量。最常见的结构模型是多元正态分布的有限混合,即

$$\sum_{c=1}^{C}\pi_c h_c(\boldsymbol{\zeta})$$

其中,$c$ 标明了成分(component),$\pi_c$ 是成分权重,$h_c(\boldsymbol{\zeta})$ 是含成分均值和协方差参数的多元正态密度。这个模型曾被沃贝克和勒斯弗(Verbeke and Lesaffre, 1996),阿伦比等(Allenby et al., 1998),以及伦克和德萨伯(Lenk and DeSarbo, 2000)用以随机系数模型;被麦格德和泽戈(Magder and Zeger, 1996),卡洛尔等(Carroll et al., 1999),以及理查森等(Richardson et al., 2002)用于协变量测量误差;被尤伯萨克斯(Uebersax, 1993),尤伯萨

克斯和格鲁夫(Uebersax and Grove,1993)用于二项和定序变量的测量模型。

第四种,即最为一般的模型允许含连续响应的传统结构方程模型的**任意**(any)参数取决于离散潜变量。因而,响应模型和结构模型两者都在潜类别之间产生差别,潜类别能提供一个此类的多组结构方程模型(multiple-group structural equation model),由杰拉斯科格提出(Jöreskog,1971a),其中关键的差别在于组隶属是否是未知的。

容(Yung,1997)、福可和蒂特林顿(Fokoué and Titterington,2003)等,一起考察了有限混合因子模型的特殊情况。容的模型可以表达为

$$\mathbf{y}_{jc} = \boldsymbol{\beta}_c + \boldsymbol{\Lambda}_c \boldsymbol{\eta}_{jc} + \boldsymbol{\epsilon}_{jc} \tag{4.17}$$

其中,$\boldsymbol{\eta}_{jc}$是连续公共因子,它的类别方差为

$$\boldsymbol{\Psi}_c \equiv \mathrm{Var}[\boldsymbol{\eta}_{jc}]$$

独特因子的类别协方差矩阵为

$$\boldsymbol{\Theta}_c \equiv \mathrm{Var}[\boldsymbol{\epsilon}_{jc}]$$

福可和蒂特林顿假设 $\boldsymbol{\Theta}_c = \boldsymbol{\Theta}$ 且 $\boldsymbol{\Psi}_c = \mathbf{I}$。在诊断测试协议(diagnostic test agreement)中,曲等(Qu et al.,1996)设定一个单维对数概率模型。他们将这个模型解释为一个有“随机效应”(公共因子)的潜类模型来释放条件独立。

Blåfield(1980)、Jedidi et al.(1997)、Dolan and van der Maas(1998)、Arminger et al.(1999)、McLachlan and Peel(2000)、Wedel and Kamakura(2000)、Muthén(2002)等,设定了“有限混合结构方程模型”,包括这样的一个结构模型:

$$\boldsymbol{\eta}_{jc} = \mathbf{B}_c \boldsymbol{\eta}_{jc} + \boldsymbol{\Gamma}_c \mathbf{w}_{jc} + \boldsymbol{\zeta}_{jc}, \quad \boldsymbol{\Psi}_c \equiv \mathrm{Var}[\boldsymbol{\zeta}_{jc}]$$

其因子见模型(4.17)

## 4.5 参数约束和基本参数

假定“数据产生模型”的参数已经产生了观测数据,这里称为**结构参数**。我们用 $\boldsymbol{\theta}$ 来表示包括回归系数 $\boldsymbol{\beta}$ 在内的所有结构参数的向量,因子负荷为 $\boldsymbol{\lambda}_m^l, m = 1,\cdots,M_l, l = 1,\cdots,L$,协方差矩阵的非重复元素为 $\boldsymbol{\Psi}^{(l)}, l = 1,\cdots,L$,用参数 $\iota$ 来对第一层次的异方差性建模,阀值参数为$\boldsymbol{\varsigma}$,类别隶属参数(*class membership parameter*)为$\boldsymbol{\varrho}$。要注意,不要将结构参数向量 $\boldsymbol{\theta}$ 与由元素 $\theta_{ii'}$所组成的残协方差矩阵 $\boldsymbol{\Theta}$ 相混淆。

通常来说,还需要进一步的参数约束,如参数的符号及某几个参数是否相等。根据无约束的所谓的**基本参数** $\boldsymbol{\vartheta}$,通过再参数化的方法能够加入约束参数。所得到的再参数化模型与原始的含有约束参数的结构模型是等价的(在第 5 章所介绍的意义上)。这种方法意在求隐函数的解,这些函数包含在有一组基本参数的结构参数之中。这种方法的优点还在于,**无约束的**最优化过程可在评估阶段使用,这就避免了含约束参数的更复杂的评估方法,如拉格朗日乘数法(Lagrange multipliers)(请参见 Fletcher,1987)。

每个结构参数 $\theta_k$设为基本参数的一个**已知**可区分时间函数,即

$$\theta_k = h_k(\boldsymbol{\vartheta})$$

可以看到,只要 $\theta_k$ 未受约束,对于某个 $l, h_k(\theta_k) = \vartheta_l$,对于大多数结构参数都如此。根据向量函数,所有约束参数数组可写为

$$\boldsymbol{\theta} = \mathbf{h}(\boldsymbol{\vartheta})$$

下面列举了不同类型的约束及其执行过程,即依据基本参数 $\boldsymbol{\vartheta}$,对结构参数 $\boldsymbol{\theta}$ 的再参数化。设有 $K$ 个结构参数 $\theta_k, k = 1,\cdots,K$,并且 $a_k, b_k$ 和 $a$ 为常数,可以为零。

(1)同一性约束(identity restriction)是可能是最常见的。例如,在增长曲线建模中,

与同方差性假设一致，残方差被约束为在不同时点下是相等的。约束采用如下形式

$$\theta_k = \theta_l, \quad k \neq l$$

其实现即为

$$\theta_k = \vartheta_k$$
$$\theta_l = \vartheta_k$$

(2)线性约束(linear restriction)采用如下形式

$$\sum_{k=1}^{K} a_k \theta_k = a$$

这样的约束对于简化模型结构非常有用。例如，对于一个定序响应阀值模型，线性约束可以用来使阈值间距相等，具体表达为

$$\theta_k = \vartheta_k; \forall k \neq K \quad 且 \quad \theta_K = \frac{a - \sum_{k=1}^{K-1} a_k \theta_k}{a_K}$$

(3)不等价约束(inequality restriction)通常采用如下形式

$$\theta_k \geqslant a_k \quad 且 \quad \theta_k > a_k$$

典型的例子如方差为正数。约束可设定为

$$\theta_k = a_k + (\vartheta_k)^2$$

且

$$\theta_k = a_k + \exp(\vartheta_k)$$

要注意的是，在无约束参数被估计小于或者等于 $a_k$ 时，约束参数必须是 $a_k$，以此来最大化似然值，第二次参数化可能在收敛时有困难。这是因为 $\vartheta_k$ 取值为较大的负值，似然值则会呈现出扁平性。

另一种类型的不等价约束是如下形式的次序约束

$$\theta_1 \leqslant \theta_2 \leqslant \cdots \leqslant \theta_K \quad 且 \quad \theta_1 < \theta_2 < \cdots < \theta_K$$

可以这样实现

$$\theta_1 = \vartheta_1, \quad \theta_k = \vartheta_{k-1} + (\vartheta_k)^2, \quad k > 1$$

且

$$\theta_1 = \vartheta_1, \quad \theta_k = \vartheta_{k-1} + \exp(\vartheta_k), \quad k > 1$$

在2.3.4节中描述的铅版模型和累积模型中需要此种约束，其中铅版模型中，$\alpha^1 < \cdots < \alpha^{S-1}$；而在累积模型中，$\kappa_{i1} < \cdots < \kappa_{iS-1}$。要注意的是，对于这些模型，即使没有明确地加入这些约束，在模型估计中也要遵循这些约束。

(4)如下形式的域约束(domain restriction)

$$a_k < \theta_k < b_k$$

可以通过实现为

$$\theta_k = \frac{a_k + b_k \exp(\vartheta_k)}{1 + \exp(\vartheta_k)}$$

例如，概率必须在[0,1]间变动，通过我们所熟悉的多项对数概率转换，对于诸如潜类概率特别有用。通过设定 $a_k = -1$ 及 $b_k = 1$，相关被约束在允许的范围内。

(5)非线性约束(nonlinear restriction)通常采用如下形式

$$\sum_{k=1}^{K} a_k \prod_{l=1}^{L} \theta_l^{b_l} = a$$

一个简单的例子是$\frac{\theta_1}{\theta_2} = \frac{\theta_3}{\theta_4}$，它可简单地转化为

$$\theta_1 = \frac{\vartheta_3}{\vartheta_4}\vartheta_2 \qquad 其中\ \theta_k = \vartheta_k; k = 2,3,4$$

它主要用来约束两个信度(真实得分与总方差的比值)一致。这个在10.4节的生活满意度案例中有所阐述,见表10.13。

非线性不等价约束隐含的要求是矩阵 $\mathbf{M}$ 是半正定矩阵(如对于所有的 $\mathbf{a}$, $\mathbf{a}'\mathbf{M}\mathbf{a} \geqslant 0$),典型的如协方差矩阵 $\mathbf{M}$,可以用**乔莱斯基分解**(Cholesky decomposition),即

$$\mathbf{M} = \mathbf{L}\mathbf{L}'$$

来增加这个约束,其中,$\boldsymbol{\theta}$ 是 $\mathbf{M}$ 的非重复元素,$\boldsymbol{\vartheta}$ 是 $\mathbf{L}$ 的下对角元素。

## 4.6 潜变量和线性估计量的简化式

### 4.6.1 $\boldsymbol{\eta}_j$ 的简化式

设 $\boldsymbol{\eta}_j$ 是 $\boldsymbol{\eta}_{jk\cdots z}$ 的简略符号,由所有第 $j$ 个第二层次单元的潜变量组成。假设 $(\mathbf{I}-\mathbf{B})$ 是可逆的,式(4.13)的结构模型为

$$\boldsymbol{\eta}_j = \mathbf{B}\boldsymbol{\eta}_j + \boldsymbol{\Gamma}\mathbf{w}_j + \boldsymbol{\zeta}_j$$

可根据探索性变量和干扰项组成的潜变量模型简化式,予以求解,即

$$\begin{aligned}\boldsymbol{\eta}_j &= (\mathbf{I}-\mathbf{B})^{-1}\boldsymbol{\Gamma}\mathbf{w}_j + (\mathbf{I}-\mathbf{B})^{-1}\boldsymbol{\zeta}_j \\ &= \boldsymbol{\Pi}_1\mathbf{w}_j + \boldsymbol{\Pi}_2\boldsymbol{\zeta}_j\end{aligned} \tag{4.18}$$

其中,$\boldsymbol{\Pi}_1 \equiv (\mathbf{I}-\mathbf{B})^{-1}\boldsymbol{\Gamma}$ 且 $\boldsymbol{\Pi}_2 \equiv (\mathbf{I}-\mathbf{B})^{-1}$ 为参数。

### 4.6.2 $\boldsymbol{\eta}^{(l)}$ 的简化式

这里要考察模型中不同层次潜变量的简化式模型。在式(4.14)中,$\mathbf{B}$ 和 $\boldsymbol{\Gamma}$ 是上对角块,因为在给定的第 $l$ 层次上变化的潜变量,对在相同或者更高层次上变化的潜变量或者观测变量回归。$\boldsymbol{\Pi}_1$ 和 $\boldsymbol{\Pi}_2$ 也是上对角块,$\boldsymbol{\Pi}_2$ 与之类似,即

$$\boldsymbol{\Pi}_1 = \begin{bmatrix} \boldsymbol{\Pi}_1^{(22)} & \boldsymbol{\Pi}_1^{(23)} & \cdots & \boldsymbol{\Pi}_1^{(2L)} \\ \mathbf{0} & \boldsymbol{\Pi}_1^{(33)} & \cdots & \boldsymbol{\Pi}_1^{(3L)} \\ \vdots & \vdots & & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\Pi}_1^{(LL)} \end{bmatrix} \tag{4.19}$$

$l=2,\cdots,L$,各个层次潜变量简化式模型可表达为

$$\boldsymbol{\eta}^{(l)} = \boldsymbol{\Pi}_1^{(ll+)}\mathbf{w}^{(l+)} + \boldsymbol{\Pi}_2^{(ll+)}\boldsymbol{\zeta}^{(l+)} \tag{4.20}$$

其中,$\mathbf{w}^{(l+)} = (\mathbf{w}^{(l)\prime},\cdots,\mathbf{w}^{(L)\prime})'$ 是 $\mathbf{w}$ 的子向量,包含所有在 $l$ 及其以上层次变动的变量,$\boldsymbol{\zeta}^{(l+)}$ 也是如此。$\boldsymbol{\Pi}_1^{(ll+)}$ 是 $\boldsymbol{\Pi}_1$ 的子矩阵,其中,行与分块 $l$ 相当,列与 $l$ 到 $L$ 分块相当,$\boldsymbol{\Pi}_2^{(ll+)}$ 与之类似。

### 4.6.3 线性估计量 $\nu_{z(L)}$ 的简化式

现在来推导一下对于第 $z$ 个 $L$ 层次的单元,其线性估计量 $\nu_{z(L)}$ 的简化式。根据式(4.6)中的“双层次表示法”,最高层级单元的线性估计量的向量可表达为

$$\boldsymbol{\nu}_{z(L)} = \mathbf{X}_{z(L)}\boldsymbol{\beta} + \boldsymbol{\Lambda}_{z(L)}\boldsymbol{\eta}_{z(L)}$$

由于 $\boldsymbol{\eta}_{z(L)} = (\boldsymbol{\eta}_{z(L)}^{(2)},\boldsymbol{\eta}_{z(L)}^{(3)},\cdots,\boldsymbol{\eta}_{z(L)}^{(L)})'$,即第 $z$ 个最高层次单元的所有潜变量(能实现

的)向量。设 $\mathbf{\Pi}^{u}_{2(L)}=(\mathbf{\Pi}^{(ll)}_{2},\cdots,\mathbf{\Pi}^{(ll)}_{2})$,这个追加列矩阵,其被追加的次数与在第 $z$ 个 $L$ 层次单元中所包含的 $l$ 层次单元个数相同,设 $\mathbf{\Pi}_{2z(L)}$ 是一个上对角块矩阵,与式(4.19)类似,但其中含 $\mathbf{\Pi}^{(ll)}_{2z(L)}$,则

$$\boldsymbol{\eta}_{z(L)}=\mathbf{\Pi}_{2z(L)}\boldsymbol{\zeta}_{(L)}$$

从而,线性估计量可表达为(省略下标 z):

$$\begin{aligned}\nu_{(L)}&=\mathbf{X}_{(L)}\beta+\mathbf{\Lambda}_{(L)}\mathbf{\Pi}_{1(L)}\mathbf{w}_{(L)}+\mathbf{\Lambda}_{(L)}\mathbf{\Pi}_{2(L)}\boldsymbol{\zeta}_{(L)}\\&=\mathbf{X}_{(L)}\beta+\mathbf{A}_{1(L)}\mathbf{w}_{(L)}+\mathbf{A}_{2(L)}\boldsymbol{\zeta}_{(L)}\end{aligned}\tag{4.21}$$

其中,$\boldsymbol{\beta},\mathbf{A}_{1(L)}\equiv\mathbf{\Lambda}_{(L)}\mathbf{\Pi}_{1(L)}$ 及 $\mathbf{A}_{2(L)}\equiv\mathbf{\Lambda}_{(L)}\mathbf{\Pi}_{2(L)}$ 是简化式线性估计量的参数。

## 4.7 潜变量的矩结构

### 4.7.1 $\boldsymbol{\eta}_j$ 的矩结构

根据式(4.18),$\boldsymbol{\eta}_j$ 的均值结构就为

$$\mathrm{E}(\boldsymbol{\eta}_j\mid\mathbf{w}_j)=\mathbf{\Pi}_1\mathbf{w}_j\tag{4.22}$$

且协方差结构为

$$\mathrm{Cov}(\boldsymbol{\eta}_j\mid\mathbf{w}_j)=\mathbf{\Pi}_2\mathrm{Cov}(\boldsymbol{\zeta}_j)\mathbf{\Pi}_2'\tag{4.23}$$

干扰项的协方差矩阵 $\mathrm{Cov}(\boldsymbol{\zeta}_j)$ 具有分块对角形式,其在第 $l$ 层次的块表示为 $\boldsymbol{\Psi}^{(l)}$。由于干扰项在不同层次中是独立的,式(4.23)中的协方差矩阵可重写为

$$\mathrm{Cov}(\boldsymbol{\eta}_j\mid\mathbf{w}_j)=\sum_{l=2}^{L}\mathbf{\Pi}_2^{(ll)}\boldsymbol{\Psi}^{(l)}\mathbf{\Pi}_2^{(ll)\prime}\tag{4.24}$$

其中,$\mathbf{\Pi}_2^{(ll)}$ 是 $\mathbf{\Pi}_2$ 的块,如同式(4.19)中的 $\mathbf{\Pi}_1$。

### 4.7.2 $\boldsymbol{\eta}^{(l)}$ 的矩结构

从式(4.20)可知,对于第 $l$ 层次的潜变量 $\boldsymbol{\eta}^{(l)}$,给定 $\mathbf{w}^{(l+)}$,则其条件均值结构就为

$$\mathrm{E}(\boldsymbol{\eta}^{(l)}\mid\mathbf{w}^{(l+)})=\mathbf{\Pi}_1^{(ll+)}\mathbf{w}^{(l+)}$$

第 $l$ 层次的协方差结构为

$$\begin{aligned}\mathrm{Cov}(\boldsymbol{\eta}^{(l)}\mid\mathbf{w}^{(l+)})&=\mathbf{\Pi}_2^{(ll+)}\mathrm{Cov}(\boldsymbol{\zeta}^{(l+)})(\mathbf{\Pi}_2^{(ll+)})'\\&=\sum_{c=l}^{L}\mathbf{\Pi}_2^{(lc)}\boldsymbol{\Psi}^{(c)}\mathbf{\Pi}_2^{(lc)\prime}\end{aligned}\tag{4.25}$$

并且 $a$ 层次和 $b$ 层次($a<b$)之间的协方差结构则为

$$\begin{aligned}\mathrm{Cov}(\boldsymbol{\eta}^{(a)},\boldsymbol{\eta}^{(b)}\mid\mathbf{w}^{(a+)})&=\mathbf{\Pi}_2^{(ab+)}\mathrm{Cov}(\boldsymbol{\zeta}^{(b+)})(\mathbf{\Pi}_2^{(bb+)})'\\&=\sum_{c=b}^{L}\mathbf{\Pi}_2^{(ac)}\boldsymbol{\Psi}^{(c)}\mathbf{\Pi}_2^{(bc)\prime}\end{aligned}\tag{4.26}$$

其中,式(4.25)和式(4.26)中的第二个等式,与协方差 $\mathrm{Cov}(\boldsymbol{\zeta}^{(l+)})$ 的分块对角形式一致。

### 4.7.3 例:三层次模型

对下面这个特殊的三层次模型进行分析有较大的启发意义。在这个例子中,结构模型可表述为

$$\begin{bmatrix}\eta_{jk}^{(2)}\\\eta_{k}^{(3)}\end{bmatrix}=\begin{bmatrix}\mathbf{B}^{(22)}&\mathbf{B}^{(23)}\\\mathbf{0}&\mathbf{B}^{(33)}\end{bmatrix}\begin{bmatrix}\boldsymbol{\eta}_{jk}^{(2)}\\\boldsymbol{\eta}_{k}^{(3)}\end{bmatrix}+\begin{bmatrix}\mathbf{\Gamma}^{(22)}&\mathbf{\Gamma}^{(23)}\\\mathbf{0}&\mathbf{\Gamma}^{(33)}\end{bmatrix}\begin{bmatrix}\mathbf{w}_{jk}^{(2)}\\\mathbf{w}_{k}^{(3)}\end{bmatrix}+\begin{bmatrix}\boldsymbol{\zeta}_{jk}^{(2)}\\\boldsymbol{\zeta}_{k}^{(3)}\end{bmatrix}$$

根据式(4.20),潜变量的简化式为

$$\boldsymbol{\eta}_{jk}^{(2)} = \boldsymbol{\Pi}_1^{(22+)}\mathbf{w}_{jk}^{(2+)} + \boldsymbol{\Pi}_2^{(22+)}\boldsymbol{\zeta}_{jk}^{(2+)}$$

$$= \begin{bmatrix}\boldsymbol{\Pi}_1^{(22)} & \boldsymbol{\Pi}_1^{(23)}\end{bmatrix}\begin{bmatrix}\mathbf{w}_{jk}^{(2)}\\ \mathbf{w}_k^{(3)}\end{bmatrix} + \begin{bmatrix}\boldsymbol{\Pi}_2^{(22)} & \boldsymbol{\Pi}_2^{(23)}\end{bmatrix}\begin{bmatrix}\boldsymbol{\zeta}_{jk}^{(2)}\\ \boldsymbol{\zeta}_k^{(3)}\end{bmatrix}$$

且

$$\boldsymbol{\eta}_k^{(3)} = \boldsymbol{\Pi}_1^{(33)}\mathbf{w}_k^{(3)} + \boldsymbol{\Pi}_2^{(33)}\boldsymbol{\zeta}_k^{(3)}$$

通过首先对 $\boldsymbol{\eta}_k^{(3)}$ 求解,然后把结构模型中的简化式 $\boldsymbol{\eta}_k^{(3)}$ 替代成 $\boldsymbol{\eta}_{jk}^{(2)}$,并对 $\boldsymbol{\eta}_{jk}^{(2)}$ 求解,能得到参数矩阵的表达式,即

$$\boldsymbol{\Pi}_1^{(22)} = (\mathbf{I} - \mathbf{B}^{(22)})^{-1}\boldsymbol{\Gamma}^{(22)}$$

$$\boldsymbol{\Pi}_1^{(23)} = (\mathbf{I} - \mathbf{B}^{(22)})^{-1}\left[\mathbf{B}^{(23)}(\mathbf{I} - \mathbf{B}^{(33)})^{-1}\boldsymbol{\Gamma}^{(33)} + \boldsymbol{\Gamma}^{(23)}\right]$$

$$\boldsymbol{\Pi}_2^{(22)} = (\mathbf{I} - \mathbf{B}^{(22)})^{-1}$$

$$\boldsymbol{\Pi}_2^{(23)} = (\mathbf{I} - \mathbf{B}^{(22)})^{-1}\mathbf{B}^{(23)}(\mathbf{I} - \mathbf{B}^{(33)})^{-1}$$

$$\boldsymbol{\Pi}_1^{(33)} = (\mathbf{I} - \mathbf{B}^{(33)})^{-1}\boldsymbol{\Gamma}^{(33)}$$

$$\boldsymbol{\Pi}_2^{(33)} = (\mathbf{I} - \mathbf{B}^{(33)})^{-1}$$

条件期望值则为

$$\begin{aligned}\mathrm{E}(\boldsymbol{\eta}_{jk}^{(2)} \mid \mathbf{w}_{jk}^{(2)}, \mathbf{w}_k^{(3)}) &= \boldsymbol{\Pi}_1^{(22)}\mathbf{w}_{jk}^{(2)}\boldsymbol{\Pi}_1^{(23)}\mathbf{w}_k^{(3)}\\ &= (\mathbf{I} - \mathbf{B}^{(22)})^{-1} \times\\ &\quad \left\{\left[\mathbf{B}^{(23)}(\mathbf{I} - \mathbf{B}^{(33)})^{-1}\boldsymbol{\Gamma}^{(33)} + \boldsymbol{\Gamma}^{(23)}\right]\mathbf{w}_k^{(3)} + \boldsymbol{\Gamma}^{(22)}\mathbf{w}_{jk}^{(2)}\right\}\end{aligned}$$

且

$$\begin{aligned}\mathrm{E}(\boldsymbol{\eta}_k^{(3)} \mid \mathbf{w}_k^{(3)}) &= \boldsymbol{\Pi}_1^{(33)}\mathbf{w}_k^{(3)}\\ &= (\mathbf{I} - \mathbf{B}^{(33)})^{-1}\boldsymbol{\Gamma}^{(33)}\mathbf{w}_k^{(3)}\end{aligned}$$

条件协方差矩阵则为

$$\begin{aligned}\mathrm{Cov}(\boldsymbol{\eta}_{jk}^{(2)} \mid \mathbf{w}_{jk}^{(2)}, \mathbf{w}_k^{(3)}) &= \boldsymbol{\Pi}_2^{(22)}\boldsymbol{\Psi}^{(2)}\boldsymbol{\Pi}_2^{(22)\prime} + \boldsymbol{\Pi}_2^{(23)}\boldsymbol{\Psi}^{(3)}\boldsymbol{\Pi}_2^{(23)\prime}\\ &= (\mathbf{I} - \mathbf{B}^{(22)})^{-1} \times\\ &\quad \left[\mathbf{B}^{(23)}(\mathbf{I} - \mathbf{B}^{(33)})^{-1}\boldsymbol{\Psi}^{(3)}(\mathbf{I} - \mathbf{B}^{(33)})^{-1\prime}\mathbf{B}^{(23)\prime} + \boldsymbol{\Psi}^{(2)}\right] \times\\ &\quad (\mathbf{I} - \mathbf{B}^{(22)})^{-1\prime}\end{aligned}$$

$$\begin{aligned}\mathrm{Cov}(\boldsymbol{\eta}_k^{(3)} \mid \mathbf{w}_k^{(3)}) &= \boldsymbol{\Pi}_2^{(33)}\boldsymbol{\Psi}^{(3)}\boldsymbol{\Pi}_2^{(33)\prime}\\ &= (\mathbf{I} - \mathbf{B}^{(33)})^{-1}\boldsymbol{\Psi}^{(3)}(\mathbf{I} - \mathbf{B}^{(33)})^{-1\prime}\end{aligned}$$

且

$$\begin{aligned}\mathrm{Cov}(\boldsymbol{\eta}_{jk}^{(2)}, \boldsymbol{\eta}_k^{(3)} \mid \mathbf{w}_{jk}^{(2)}, \mathbf{w}_k^{(3)}) &= \boldsymbol{\Pi}_2^{(23)}\boldsymbol{\Psi}^{(3)}\boldsymbol{\Pi}_2^{(33)\prime}\\ &= (\mathbf{I} - \mathbf{B}^{(22)})^{-1}\mathbf{B}^{(23)}(\mathbf{I} - \mathbf{B}^{(33)})^{-1}\boldsymbol{\Psi}^{(3)}(\mathbf{I} - \mathbf{B}^{(33)})^{-1\prime}\end{aligned}$$

## 4.8 观测响应和潜响应的边缘矩结构

根据响应模型条件依赖于潜变量,响应模型就是一个含链接函数 $g(\nu)$ 的广义线性模型。因而,回归参数则表示给定潜变量条件下的协变量的**条件效应**。在某些情况下,

**总体平均效应**(population averaged effects)或者边缘效应(对潜变量分布平均)是值得我们关注的。首先考察响应的均值结构,将响应的期望值作为协变量的函数但边缘于潜变量。其次再推导出边缘协方差结构。

### 4.8.1 边缘均值结构和总体平均效应

现在考察一下条件依赖于协变量,但边缘于潜变量的响应的期望值,应用双重期望规则(double expectation rule),即

$$\mathrm{E}(\mathbf{y}_{(L)} \mid \mathbf{X}_{(L)}, \mathbf{w}_{(L)}) = \mathrm{E}_{\zeta}\left\{\mathrm{E}(\mathrm{y}_{(L)} \mid \mathbf{X}_{(L)}, \mathbf{w}_{(L)}, \boldsymbol{\zeta}_{(L)})\right\}$$

$$= \int\cdots\int g^{-1}(\boldsymbol{\nu}_{(L)})\left[\prod_{l=2}^{L} h^{(l)}(\boldsymbol{\zeta}_{(L)}^{(l)})\right]\mathrm{d}\boldsymbol{\zeta}_{(L)}^{(2)}\cdots\mathrm{d}\boldsymbol{\zeta}_{(L)}^{(L)} \qquad (4.27)$$

其中,$h^{(l)}(\cdot)$是在第 $l$ 层次的干扰项的密度。这仅仅是一个给定协变量值的"总体平均"响应。所谓的**总体平均效应**或者**边缘效应**(对于潜变量),要通过考察期望值和协变量的关系获得。

考察式(4.21)中的线性估计量的一般形式,只有一个第一层次单元时,公式可写为

$$\nu = \mathbf{x}'\beta + \mathbf{a}_1'\mathbf{w}_{(L)} + \mathbf{a}_2'\boldsymbol{\zeta}_{(L)}$$

其中,$\mathbf{a}_1'$是 $\mathbf{A}_{1(L)}$ 的其中一行,$\mathbf{a}_2'$也是如此。例如,在一个包含连续潜变量的双层次随机系数模型中,则

$$\nu_{ij} = \mathbf{x}_{ij}'\beta + \mathbf{z}_{ij}'\boldsymbol{\zeta}_j^{(2)}$$

$\mathbf{a}_1 = \mathbf{0}, \mathbf{a}_2 = \mathbf{z}_{ij}$并且$\boldsymbol{\zeta}_{(L)} = \boldsymbol{\zeta}_j^{(2)}$。因子模型也有相同的结构,唯一的差别是,$\mathbf{a}_2'$是因子负荷矩阵 $\mathbf{\Lambda}$ 的其中一行。

对同一性链接而言,$g^{-1}(\nu) = \nu$,期望值可简化为

$$\mathrm{E}(y \mid \mathbf{x}, \mathbf{w}_{(L)}) = \mathbf{x}'\beta + \mathbf{a}_1'\mathbf{w}_{(L)} \qquad (4.28)$$

这样,链接函数得以保留,且边缘效应等于条件效应,无论潜变量是何种分布。

对于对数链接来说,$g^{-1}(\nu) = \exp(\nu)$,与潜变量分布无关,除了截距外,边缘效应等于条件效应。对于正态潜变量,可得

$$\mathrm{E}(y \mid \mathbf{x}, \mathbf{w}_{(L)}) = \exp(\mathbf{x}'\beta + \mathbf{a}_1'\mathbf{w}_{(L)} + \mathbf{a}_2'\mathbf{\Psi}_{(L)}\mathbf{a}_2/2)$$

此式在随机截距的情形下,可简化为

$$\mathrm{E}(y_{ij} \mid \mathbf{x}_{ij}, \mathbf{z}_{ij}) = \exp(\mathbf{x}_{ij}'\beta + \psi/2)$$

这样,除了截距增长了 $\psi/2$ 以外,边缘回归参数仍然等于条件参数。

之后,解则被限定在潜变量范围内。对于一个正态概率链接(probit link)来说,用潜响应方程(见 2.4 节)推导期望值是比较简便的做法,模型可表达为

$$y^* = \mathbf{x}'\beta + \mathbf{a}_1'\mathbf{w}_{(L)} + \underbrace{\mathbf{a}_2'\boldsymbol{\zeta}_{(L)} + \epsilon}_{\xi} \qquad (4.29)$$

其中,$\xi$ 是"总残差",它是均值为零,方差为 $\mathbf{a}_2'\mathbf{\Psi}_{(L)}\mathbf{a}_2 + 1$ 的正态分布,另外,当 $y^* > 0$ 时 $y = 1$,否则 $y = 0$。请注意,潜响应的均值结构与 4.28 节同一性链接中的一致。

观测响应的期望值则变为

$$\mathrm{E}(y \mid \mathbf{x}, \mathbf{w}) = \Pr(y = 1 \mid \mathbf{x}, \mathbf{w}) = \Pr(y^* > 0 \mid \mathbf{x}, \mathbf{w})$$

$$= \Pr(-\xi \leqslant \mathbf{x}'\boldsymbol{\beta} + \mathbf{a}_1'\mathbf{w}_{(L)}) = \Pr(\xi \leqslant \mathbf{x}'\boldsymbol{\beta} + \mathbf{a}_1'\mathbf{w}_{(L)})$$

$$= \Pr\left(\frac{\xi}{\sqrt{\mathbf{a}_2'\mathbf{\Psi}_{(L)}\mathbf{a}_2 + 1}} \leqslant \frac{\mathbf{x}'\boldsymbol{\beta} + \mathbf{a}_1'\mathbf{w}_{(L)}}{\sqrt{\mathbf{a}_2'\mathbf{\Psi}_{(L)}\mathbf{a}_2 + 1}}\right)$$

$$= \Phi\left(\frac{\mathbf{x}'\boldsymbol{\beta} + \mathbf{a}_1'\mathbf{w}_{(L)}}{\sqrt{\mathbf{a}_2'\boldsymbol{\Psi}_{(L)}\mathbf{a}_2 + 1}}\right) = \Phi(\mathbf{x}'\beta^* + \mathbf{a}_1^{*\prime}\mathbf{w}_{(L)}) \qquad (4.30)$$

其中,$\boldsymbol{\beta}^*$ 和 $\mathbf{a}_1^*$ 是初始参数向量,除以潜响应的总残差的标准差。对于正态概率链接,链接函数则得以保留,但相对于条件效应,边缘效应则衰减了。基于同样的推导,对定序响应的累积正态概率模型来说,同样的结果也成立。

有时在正态概率模型中,潜响应 $y^*$ 的方差被设定为 1,$\mathbf{a}_2'\boldsymbol{\Psi}_{(L)}\mathbf{a}_2 + \theta = 1$,而非将误差项 $\epsilon$ 的方差如上述设定;比如穆森(Muthén,1984)。注意在此例中,边缘效应并未衰减。

对于对数概率链接来说,模型同样如式(4.29),只是将 $\epsilon$ 设定为对数函数,得

$$\mathrm{Var}(\xi) = \mathbf{a}_2'\boldsymbol{\Psi}_{(L)}\mathbf{a}_2 + \pi^2/3$$

总残差为复合对数概率正态分布(compound logistic-normal distribution),总残差乘以因子

$$\sqrt{\frac{\pi^2/3}{\mathbf{a}_2'\boldsymbol{\Psi}_{(L)}\mathbf{a}_2 + \pi^2/3}} \approx \frac{1}{\sqrt{1 + 0.30\mathbf{a}_2'\boldsymbol{\Psi}_{(L)}\mathbf{a}_2}}$$

能得到一个方差为对数概率分布的残差项。对数概率分布与正态分布非常类似,即

$$\Pr\left(\frac{\xi}{\sqrt{1 + 0.30\mathbf{a}_2'\boldsymbol{\Psi}_{(L)}\mathbf{a}_2}} \leqslant \frac{\mathbf{x}'\beta + \mathbf{a}_1'\mathbf{w}_{(L)}}{\sqrt{1 + 0.30\mathbf{a}_2'\boldsymbol{\Psi}_{(L)}\mathbf{a}_2}}\right) \approx F\left(\frac{\mathbf{x}'\beta + \mathbf{a}_1'\mathbf{w}_{(L)}}{\sqrt{1 + 0.30\mathbf{a}_2'\boldsymbol{\Psi}_{(L)}\mathbf{a}_2}}\right)$$

其中,$F$ 是对数概率累积分布函数。泽戈等(Zeger,1988)用 $\frac{(15/16)^2}{\pi^2/3} \approx 0.35$ 代替了 0.30。

一般来说,其他的链接函数并没有在边缘化下得到保留。瑞斯和斯皮格曼(Rith and Spiegelman,2004)的关于不同模型的边缘或者条件效应的论述很有意义。

### 4.8.2 边缘协方差结构

一般而言,边缘于潜变量的协方差结构可用下面的这一关系推导出来,即

$$\mathrm{Cov}(\mathbf{y}_{(L)} \mid \mathbf{X}_{(L)}, \mathbf{w}_{(L)}) = \mathrm{E}_{\zeta}\left[\mathrm{Cov}(\mathbf{y}_{(L)} \mid \boldsymbol{\nu}_{(L)})\right] + \mathrm{Cov}_{\zeta}\left[\mathrm{E}(\mathbf{y}_{(L)} \mid \boldsymbol{\nu}_{(L)})\right]$$

第一层次的不同单元 $i$ 和 $i'$(省略了较高层次的下标)之间的边缘协方差则为

$$\mathrm{Cov}(y_i, y_{i'} \mid \mathbf{X}_{(L)}, \mathbf{w}_{(L)}) = \mathrm{E}_{\zeta}\left[\mathrm{Cov}(y_i, y_{i'} \mid \nu_i, \nu_{i'})\right] + \mathrm{Cov}_{\zeta}\left[\mathrm{E}(y_i \mid \nu_i), \mathrm{E}(y_{i'} \mid \nu_{i'})\right]$$

$$= \mathrm{Cov}_{\zeta}\left[g^{-1}(\nu_i), g^{-1}(\nu_{i'})\right] \qquad (4.31)$$

即为条件期望之间的协方差。要注意式(4.31)的第 1 项是 0,因为如果是潜变量条件独立的话,则 $\mathrm{Cov}(y_i, y_{i'} | \nu_i, \nu_{i'}) = 0$。具有两个相同线性估计量值的单元对,将有同样的边缘协方差和相关。尽管如此,同一性链接除外,如果存在较低层次的协变量,则观测响应的"组内"相关,在具有不同协变量的群集之间将会不同,在同一群集内的不同单元对之间也会不同(如纵贯性或者面板数据中的时间)。

对于连续响应和潜响应,边缘协方差结构为

$$\boldsymbol{\Omega}_{(L)} \equiv \mathrm{Cov}(\mathbf{y}_{(L)}^* \mid \mathbf{X}_{(L)}, \mathbf{w}_{(L)}) = \mathbf{A}_{2(L)}\boldsymbol{\Psi}_{(L)}\mathbf{A}_{2(L)}' + \boldsymbol{\Theta} \qquad (4.32)$$

其中,$\boldsymbol{\Theta}$ 是"误差"$\epsilon$ 的典型的对角协方差矩阵(见展示 4.1)。这样即可得到相关结构为

$$\rho_{(L)} \equiv \mathrm{Cor}(\mathbf{y}_{(L)}^* \mid \mathbf{X}_{(L)}, \mathbf{w}_{(L)}) = [\mathrm{Diag}(\boldsymbol{\Omega}_{(L)})]^{-\frac{1}{2}}\boldsymbol{\Omega}_{(L)}[\mathrm{Diag}(\boldsymbol{\Omega}_{(L)})]^{-\frac{1}{2}} \qquad (4.33)$$

考察一下多元正态潜响应这样一个特例。响应类型和相应的潜响应的相关的名称见表 4.2。过去,这些称谓指二元变量的相关,没有条件依赖于解释变量,但这里将一般

化来使用。

表 4.2 双变量正态潜响应的相关性命名法则

| | 连 续 | 二 项 | 定 序 | 删 失 |
|---|---|---|---|---|
| 连续 | 皮尔逊 | | | |
| 二项 | 二列 | 四项 | | |
| 定序 | 多序列 | 多项 | 多项 | |
| 删失 | Tobitserial? | Bitobit? | Polytobit? | 托比(Tobit) |

注:我们自己创造了这些词[1]。

对于包含对数链接的泊松响应来讲,其边缘方差为

$$
\begin{aligned}
\mathrm{Var}(y \mid \mathbf{x},\mathbf{w}_{(L)}) &= \mathrm{E}_{\zeta}\left[\mathrm{Var}(y \mid \mathbf{x},\mathbf{w}_{(L)},\boldsymbol{\zeta}_{L})\right] + \mathrm{Var}_{\zeta}\left[\mathrm{E}(y \mid \mathbf{x},\mathbf{w}_{(L)},\zeta_{(L)})\right] \\
&= \mathrm{E}_{\zeta}\left[\exp(\mathbf{x}'\boldsymbol{\beta} + \mathbf{a}_1'\mathbf{w}_{(L)} + \mathbf{a}_2'\boldsymbol{\zeta}_{(L)})\right] + \mathrm{Var}_{\zeta}\left[\exp(\mathbf{x}'\boldsymbol{\beta} + \mathbf{a}_1'\mathbf{w}_{(L)} + \mathbf{a}_2'\boldsymbol{\zeta}_{(L)})\right] \\
&= \mathrm{E}(y \mid \mathbf{x},\mathbf{w}_{(L)}) + \exp(\mathbf{x}'\boldsymbol{\beta} + \mathbf{a}_1'\mathbf{w}_{(L)})^2\mathrm{Var}\left[\exp(\mathbf{a}_2'\boldsymbol{\zeta}_{(L)})\right]
\end{aligned}
$$

对于正态$\boldsymbol{\zeta}$,根据对数-正态随机变量的方差的一般求解(请参见 Johnson et al.,1994:212),即

$$
\mathrm{Var}(\exp(\mathbf{a}_2'\boldsymbol{\zeta}_L)) = \exp(\mathbf{a}_2'\boldsymbol{\Psi}_{(L)}\mathbf{a}_2)\left[\exp(\mathbf{a}_2'(\boldsymbol{\Psi})_{(L)}\mathbf{a}_2) - 1\right]
$$

故

$$
\mathrm{Var}(y \mid \mathbf{x},\mathbf{w}_{(L)}) = \mathrm{E}(y \mid \mathbf{x},\mathbf{w}_{(L)})\left\{1 + \mathrm{E}(y \mid \mathbf{x},\mathbf{w}_{(L)})\left[\exp(\mathbf{a}_2'\boldsymbol{\Psi}_{(L)}\mathbf{a}_2) - 1\right]\right\}
$$

若没有潜变量,边缘方差则化约为 $\mathrm{E}(y \mid \mathbf{x},\mathbf{w}_{(L)})$,即泊松分布的方差函数。因而,潜变量使方差的增大,与期望值的平方成正比。相反,通常的准似然方法将方差函数设为等于 $\phi^{*}\mathrm{E}(y \mid \mathbf{x},\mathbf{w}_{(L)})$(见 2.3.1 节),且假设方差与期望值成正比。

对二项响应来说,边缘方差为

$$
\mathrm{Var}(y \mid \mathbf{x},\mathbf{w}_{(L)}) = \Pr(y = 1 \mid \mathbf{x},\mathbf{w}_{(L)})\left[1 - \Pr(y = 1 \mid \mathbf{x},\mathbf{w}_{(L)})\right]
$$

如果是正态概率链接,则可简化为

$$
\Phi\left(\frac{\mathbf{x}'\boldsymbol{\beta} + \mathbf{a}_1'\mathbf{w}_{(L)}}{\sqrt{\mathbf{a}_2'\boldsymbol{\Psi}_{(L)}\mathbf{a}_2 + 1}}\right)\left[1 - \Phi\left(\frac{\mathbf{x}'\boldsymbol{\beta} + \mathbf{a}_1'\mathbf{w}_{(L)}}{\sqrt{\mathbf{a}_2'\boldsymbol{\Psi}_{(L)}\mathbf{a}_2 + 1}}\right)\right]
$$

请注意均值和方差之间的关系与在不含潜变量的伯努利模型中是相同的。这点也正如我们所期望的,因为对于二项响应来说,过度的离散是不可能的。

## 4.9 简化式分布和似然

给定解释变量 $\mathbf{X}$,观测响应 $\mathbf{y}$ 的条件分布称为**简化式分布**。通过潜变量积分或者潜响应积分,有两种方法来推导这种简化式:第一种方法是建立在给定潜变量的条件下,将观测响应设定为条件独立的基础上的;第二种方法在于设定潜响应为多元正态性,且边缘于潜变量。

1 由于作者创造了这些术语,目前未见于其他文献,暂且采用原文的形式。——译者注

### 4.9.1 潜变量积分

简化式分布是指响应的分布边缘于潜变量，但同时条件依赖于解释变量。如果潜变量是离散的，可通过在给定有限混合的条件下，对响应和类别隶属的联合概率进行加总而得到。如果潜变量是连续的，则通过有限混合积出潜变量。因而，潜变量分布通常被称为**混合分布**（mixing distribution）。在对不同层次的潜变量求积分时，要在给定更高层次的所有潜变量的条件下，利用某个层次所有单元的条件独立。

**双层次随机截距模型**

为简化起见，双层次随机截距模型为

$$\nu_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \zeta_j^{(2)}$$

响应值 $y_{ij}$ 的条件概率（或者概率密度）用 $g^{(1)}(y_{ij} \mid \mathbf{x}_{ij}, \zeta_j^{(2)}; \boldsymbol{\vartheta})$ 表示，其中，$\boldsymbol{\vartheta}$ 是基本参数的向量。给定第 $j$ 个第二层次单元的协变量矩阵 $\mathbf{X}_{j(2)}$，则该单元的响应 $\mathbf{y}_j^{(2)}$ 的边缘分布为

$$g^{(2)}(\mathbf{y}_j^{(2)} \mid \mathbf{X}_{j(2)}; \boldsymbol{\vartheta}) = \int h(\zeta_j^{(2)}) \prod_{i=1}^{n_j} g^{(1)}(y_{ij} \mid \mathbf{x}_{ij}, \zeta_j^{(2)}; \boldsymbol{\vartheta}) \mathrm{d}\zeta_j^{(2)}$$

其中，$h(\cdot)$ 是随机截距的密度，并且乘积是对处于第 $j$ 个第二层次单元之中的所有第一层次单元 $i$。由于给定随机截距，响应是条件独立的，乘积表示给定随机截距条件下响应的联合概率（密度）。

**一般模型**

采用 GRC 方程的标记，将第一层次单元 $i$ 的条件分布表示为 $g^{(1)}(y_{(1)} \mid \mathbf{X}_{(1)}, \boldsymbol{\zeta}^{(2+)}; \boldsymbol{\vartheta})$，其中，$\boldsymbol{\zeta}^{(l+)}$ 是 $l$ 及以上层次干扰项的向量。第 $l$ 层次的潜变量的多元分布表示为 $h^{(l)}(\boldsymbol{\zeta}^{(l)})$。对于某个 $l$ 层次单元，其响应的条件分布条件依赖于 $l+1$ 及以上层次的潜变量，是该单元之内 $l-1$ 层次单元的分布的函数，即

$$g^{(l)}(\mathbf{y}_{(l)} \mid \mathbf{X}_{(l)}; \boldsymbol{\zeta}^{([l+1]+)}; \boldsymbol{\vartheta}) = \int h^{(l)}(\boldsymbol{\zeta}^{(l)}) \prod g^{(l-1)}(\mathbf{y}_{(l-1)} \mid \mathbf{X}_{(l-1)}, \boldsymbol{\zeta}^{(l+)}; \boldsymbol{\vartheta}) \mathrm{d}\boldsymbol{\zeta}^{(l)}$$

当 $l$ 从 2 增加到 $L-1$，则可用这种递归关系来构建似然。因此，$L$ 层次单元的响应的简化式分布为

$$g^{(L)}(\mathbf{y}_{(L)} \mid \mathbf{X}_{(L)}; \boldsymbol{\vartheta}) = \int h^{(L)}(\zeta^{(L)}) \prod g^{(L-1)}(\mathbf{y}_{(L-1)} \mid \mathbf{X}_{(L-1)}, \boldsymbol{\zeta}^{(L)}; \boldsymbol{\vartheta}) \mathrm{d}\boldsymbol{\zeta}^{(L)}$$

所有响应的简化式分布则为乘积

$$g(\mathbf{y} \mid \mathbf{X}; \boldsymbol{\vartheta}) = \prod g^{(L)}(\mathbf{y}_{(L)} \mid \mathbf{X}_{(L)}; \vartheta) \tag{4.34}$$

采用高斯-埃尔米特积分法和其他方法的潜变量积分将在 6.3 中进行讨论。

### 4.9.2 潜响应积分

假设潜响应（边缘于潜干扰项 $\boldsymbol{\zeta}$）具有多元正态分布。由此，一元变量、二元变量等潜响应分布（边缘于其他的潜响应 $y^*$）都是正态的。

现在，考察一下隐含在第 $i$ 个观测响应 $y_i$ 下的潜响应 $y_i^*$。例如，$y_i$ 表示一个测量模型中第 $i$ 个题器（一个受试者）的响应，或者表示一个随机效应模型中第 $i$ 个单元（在一个群组中）的响应。将这个潜响应模型表示为（省略了较高层次下标）

$$y_i^* = \mu_i + \xi_i$$

其中，均值 $\mu_i$ 见式（4.28），总残差 $\boldsymbol{\xi}_{(L)}$ 的向量的协方差矩阵 $\boldsymbol{\Omega}_{(L)}$ 协方差矩阵在式（4.32）中，其对应的相关矩阵 $\boldsymbol{\rho}_{(L)}$ 在式（4.33）中。

**一元观测响应分布**

对于一个定序或者二项响应 $y_i$ 来说，给定解释变量，响应类别的边缘概率可表示为

$$\Pr(y_i = a_s) = \frac{1}{\sqrt{\omega_{ii}}}\int_{\kappa_{s-1}/\sqrt{\omega_{ii}}}^{\kappa_s/\sqrt{\omega_{ii}}} \phi\left(\frac{\mu_i + \xi_i}{\sqrt{\omega_{ii}}}\right)\mathrm{d}\xi_i \tag{4.35}$$

其中，$\phi$ 是标准正态密度，$\omega_{ii}$是 $\boldsymbol{\Omega}_{(L)}$ 的第 $i$ 个对角元素。积分限则为**简化式阈值**，即

$$\tau_{is} = \frac{\kappa_s}{\sqrt{\omega_{ii}}}$$

对一个左侧删失的连续响应来说，其概率如上式所示，且满足 $\tau_{i,s-1} = -\infty$，$\kappa_s$ 等于删失边界。这里，一直让均值结构与阀值结构相分离；均值结构成为线性估计量的一部分，该估计量响应的不同类别中是常数。另外一种可能是将积分写为

$$\begin{aligned}\Pr(y_i = a_s) &= \frac{1}{\sqrt{\omega_{ii}}}\int_{\tau_{i,s-1}^*}^{\tau_{is}^*} \phi\left(\frac{\xi_i}{\sqrt{\omega_{ii}}}\right)\mathrm{d}\xi_i \\ &= \Phi(\tau_{is}^*) - \Phi(\tau_{is-1}^*)\end{aligned}$$

其中

$$\tau_{is}^* = \frac{\kappa_s - \mu_i}{\sqrt{\omega_{ii}}}$$

**二元观测响应分布**

对于两个定序或者二项响应变量 $y_i$ 和 $y_{i'}$，分别具有 $S$ 和 $T$ 个类别，联合响应概率为

$$\Pr(y_i = a_s, y_{i'} = b_t) = \frac{1}{\sqrt{\omega_{ii}\omega_{i'i'}}}\int_{\tau_{s-1}}^{\tau_s}\int_{\tau_{t-1}}^{\tau_t} \varphi\left(\frac{\mu_i + \xi_i}{\sqrt{\omega_{ii}}}, \frac{\mu_{i'} + \xi_{i'}}{\sqrt{\omega_{i'i'}}}; \rho_{ii'}\right)\mathrm{d}\xi_i\mathrm{d}\xi_{i'}$$

其中，$\varphi(.\,,.\,;\rho_{ii'})$是二元标准正态密度，潜响应 $y_i^*$ 和 $y_{i'}^*$ 的相关系数为 $\rho_{ii'}$。

对于删失响应，左侧删失响应（在 $\kappa_{is}$）和观测到的连续响应的响应分布为

$$\Pr(y_i^* < \kappa_{is}, y_{i'}^* = y_{i'}) = \frac{1}{\sqrt{\omega_{ii}\omega_{i'i'}}}\int_{-\infty}^{\kappa_{is}} \varphi\left(\frac{\mu_i + \xi_i}{\sqrt{\omega_{ii}}}, \frac{\mu_j + \xi_{i'}}{\sqrt{\omega_{i'i'}}}; \rho_{ii'}\right)\mathrm{d}\xi_i$$

所有二项响应、定序响应、类别响应及删失响应之间所组合的二元（以及三元等）分布都包含类似的积分。

一元和二元响应分布形成了将在 6.7 节中论述的有限信息估计方法的基础。对某个 $L$ 层次单元的所有潜响应求积分，则得到该单元的简化式分布 $g^{(L)}(\mathbf{y}^{(L)} \mid \mathbf{X}^{(L)};\boldsymbol{\vartheta})$。对高维潜变量积分的一种比较普遍的方法是模拟和仿真，见 6.3.4 节。

### 4.9.3 似 然

边缘似然$f(\boldsymbol{\vartheta};\mathbf{y}\mid\mathbf{X})$与所有单元的简化式分布成比例，并且被视为给定响应值的参数的一个函数，即

$$f(\boldsymbol{\vartheta};\mathbf{y} \mid \mathbf{X}) \propto g(\mathbf{y} \mid \mathbf{X};\boldsymbol{\vartheta}) = \prod g^{(L)}(\mathbf{y}^{(L)} \mid \mathbf{X}^{(L)};\boldsymbol{\vartheta})$$

## 4.10 简化式参数

潜变量简化式和线性估计量简化式的参数可视为简化式参数。尽管如此，这个短语常用来表述具有下列属性的参数：

(1)简化式参数是基本参数 $\boldsymbol{\vartheta}$ 的函数。

(2)简化式参数完全表现简化式分布。

(3)简化式分布仅仅通过简化式参数而依赖于基本参数。

在多元正态潜变量和条件正态观测响应的特殊例子中,最高层次单元的简化式分布是多元正态的。在此例中,分布完全由一阶矩和二阶矩(均值结构和协方差结构)所表现。由此可知,在上述属性中,"简化式分布"可以用"一阶矩和二阶矩"代替。在4.9.2节中,我们证明了(含多元正态潜变量的)正态累积概率模型的简化式分布完全由均值结构、阈值结构和四项相关所表现。因而,这些特质可代替在上面属性列表中的"简化式分布"。

为了更直观地说明上面的观点,以条件正态观测响应为例,我们现在推导一个含有4个题器的单因子模型的简化式参数:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \end{bmatrix} \eta + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \end{bmatrix}$$

其中 $\eta \sim \mathrm{N}(0,\psi)$,$\epsilon_i \sim \mathrm{N}(0,\theta_{ii})$以及 $\mathrm{Cov}(\epsilon_i,\epsilon_{i'})=0$(由于响应已进行了均值中心化,故没有设定截距)。总之,列向量 $\boldsymbol{\vartheta}$ 中共有9个未知参数,即

$$\boldsymbol{\vartheta} = \left[\lambda_1,\lambda_2,\lambda_3,\lambda_4,\psi,\theta_{11},\theta_{22},\theta_{33},\theta_{44}\right]' \tag{4.36}$$

由于模型暗含了 $\mathrm{E}(y_i)=0$ 以及正态性设定,可知所有信息都包含在 $y_i$ 的二阶矩,即

$$\boldsymbol{\Omega}(\boldsymbol{\vartheta}) \equiv \mathrm{Cov}(\mathbf{y}) = \begin{bmatrix} \lambda_1^2\psi+\theta_{11} & & & \\ \lambda_2\psi\lambda_1 & \lambda_2^2\psi+\theta_{22} & & \\ \lambda_3\psi\lambda_1 & \lambda_3\psi\lambda_2 & \lambda_3^2\psi+\theta_{33} & \\ \lambda_4\psi\lambda_1 & \lambda_4\psi\lambda_2 & \lambda_4\psi\lambda_3 & \lambda_4^2\psi+\theta_{44} \end{bmatrix}$$

总共有10个非冗余的协变量和协方差,都处于简化式参数 $\mathbf{m}(\boldsymbol{\vartheta})$ 的向量中,即

$$\begin{aligned} \mathbf{m}(\boldsymbol{\vartheta}) &= \mathrm{vech}(\boldsymbol{\Omega}(\boldsymbol{\vartheta})) \\ &= \left[\lambda_1^2\psi+\theta_{11},\lambda_2\psi\lambda_1,\lambda_3\psi\lambda_1,\lambda_4\psi\lambda_1,\lambda_2^2\psi+\theta_{22},\lambda_3\psi\lambda_2,\lambda_4\psi\lambda_2,\right. \\ &\qquad \left.\lambda_3^2\psi+\theta_{33},\lambda_4\psi\lambda_3,\lambda_4^2\psi+\theta_{44}\right] \end{aligned} \tag{4.37}$$

一般来说,情况可能更加复杂,因为协方差矩阵的元素可能会依赖于协变量,并被写成多项式的方式。在此例中,所对应的简化式参数是这些多项式的独特系数集合(直到乘积常数)。例如,对于包含一个协变量 $x_{ij}$的随机系数模型

$$y_{ij} = \beta_1 + \beta_2 x_{ij} + \eta_{1j} + \eta_{2j}x_{ij} + \epsilon_{ij}$$

其随机斜率为 $\eta_{2j}$,并对随机截距 $\eta_{1j}$回归,即

$$\eta_{2j} = b_{21}\eta_{1j} + \zeta_{2j}$$

$$\eta_{1j} = \zeta_{1j}$$

其中,$\begin{bmatrix} \zeta_{1j} \\ \zeta_{2j} \end{bmatrix} \sim \mathrm{N}_2\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix},\begin{bmatrix} \psi_{11} & \\ \psi_{21} & \psi_{22} \end{bmatrix}\right)$且 $\epsilon_{ij} \sim \mathrm{N}(0,\theta)$。则响应的条件方差为

$$\mathrm{Var}[y_{ij} \mid x_{ij}] = \psi_{11} + 2b_{21}\psi_{11}x_{ij} + (b_{21}^2\psi_{11}+\psi_{22})x_{ij}^2 + \theta$$

协方差为

$$\mathrm{Cov}[y_{ij},y_{i'j} \mid x_{i'j},x_{i'j}] = \psi_{11} + b_{21}\psi_{11}(x_{ij} + x_{x'j}) + (b_{21}^2\psi_{11} + \psi_{22})x_{ij}x_{x'j}$$

如果 $i=1,2$，简化式参数为 $\psi_{11}+\theta, b_{21}\psi_{11}, b_{21}^2\psi_{11}+\psi_{22}$（当 $i=1$ 时，方差的简化式参数）；$\psi_{11}+\theta, b_{21}\psi_{11}, b_{21}^2\psi_{11}+\psi_{22}$（当 $i=2$ 时，方差的简化式参数）；$\psi_{11}, b_{21}\psi_{11}, b_{21}^2\psi_{11}+\psi_{22}$（当 $i=1$ 和 $i=2$ 时，协方差的简化式参数）；$\beta_1,\beta_2$（当 $i=1$ 时，均值的简化式参数）；$\beta_1,\beta_2$（当 $i=2$ 时，均值的简化式参数）。这 6 个非冗余简化式参数可汇聚为

$$\mathbf{m}(\boldsymbol{\vartheta}) = \left[\beta_1,\beta_2,\psi_{11},\psi_{11}+\theta,b_{21}\psi_{11},b_{21}^2\psi_{11}+\psi_{22}\right]$$

## 4.11 总结与进一步阅读

前文已介绍了统一多层次模型、结构方程模型、潜类模型及纵贯性模型的一般模型框架。这个框架结构可容纳不同层次的因子和随机系数，各种类型的响应过程及其回归结构，以及灵活的潜变量分布设定。它实质上是拉贝-赫斯基思等(Rabe-Hesketh,2004a)在连续潜变量的例子中讨论过的广义线性潜混合模型(GLLAMM)框架。

GLLAMM 框架的发展是与 Stata 程序 **gllamm**(在www. gllamm. org 中可找到)的发展是同步的。这个程序可估计在本章中论述过的所有模型，但(到目前为止)同时包含离散和连续潜变量的模型，以及比式(4. 15)包含更多的一般结构模型的离散潜变量模型除外。进而多元正态分布是目前唯一能使用连续潜变量分布。拉贝-赫斯基思等(Rabe-Hesketh,2004c)曾经阐述过用 gllamm 进行模型估计的细节，对模型框架提出一个相对来说不太技术性的处理方法。除非有额外说明，在应用研究部分所介绍的所有模型都用 gllamm 进行估计。

其他的一些或多或少的含潜变量的一般模型框架都已被提出来了，这里只提到了几个。穆森的一般模型框架(如 Muthén,2001,2002)包括了潜特质以及潜类，能够处理连续响应、二项响应和定序响应。其他的几位对采用连续响应的多层次结构方程建模作出开创性贡献的还有 Goldstein and McDonald(1988)、McDonald and Goldstein(1989)。Bryk(2002)、Hox(2002)以及 de Boeck、Wilson(2004)等对此也有过简短的论述。福克斯(Fox ,2001)从贝叶斯定理的角度考察了多层次题器响应模型。斯科隆多(Skrondal,1996)考察了包含连续响应、删失响应、二项响应及定序响应的潜特质模型和多层次模型。在单层次设定中，塞缪等(1997)、穆斯塔奇和诺特(Moustaki and Knott,2000)论述了包含连续响应、二项响应及定序响应的潜特质模型。巴托罗买和诺特(Bartholomew and Knott,1999)以及穆斯塔奇(Moustaki,1996)同时论述了连续响应、二项响应及多项响应进程的潜特质和潜类模型。阿明戈尔和库斯特(Arminger and Küster,1988,1989)论述了连续响应、二项响应、定序响应、多项响应及事件数模型。哈格纳斯(Hagenaars,1993)和沃芒(Vermunt,1997)涉及结构方程的潜类模型。

我们使用了两种方法推导简化式分布：潜变量积分和潜响应积分。潜变量积分的一个优势是，它能处理所有的响应类型。而其劣势则是给定潜变量条件下，必须设定响应的条件依赖。尽管如此，通过使用额外潜变量引入依赖的方法能够部分克服这个缺点。潜响应积分的优点在于能够很容易地释放条件独立假设，但其缺点是只适用于包含潜响应公式的响应模型，而不适用于诸如含有对数概率链接或者泊松分布的模型。

# 5 辨识与等价

## 5.1 介　绍

本书所分析的统计模型非常复杂,远远超过仅仅对数据的描述或者探索性分析。基本的想法在于建立结构化的统计模型,以此来表示"数据生成机制"(即由经验过程所产生的观测数据)的主要特性。在这种情形下,统计模型的辨识及等价成为非常重要的问题,因为在科学上所感兴趣的结构参数,通常不同于简化式参数。一般来说,辨识与等价关注如何从基于简化式分布的结构模型中对观测到的变量进行推论。

当存在参数向量 $\boldsymbol{\vartheta}$ 的唯一值或**参数点**,使得可产生所给定的简化式分布 $g(\mathbf{y}|\mathbf{X},\boldsymbol{\vartheta})$ 时,这个参数统计模型则被称为可辨识的。如果一个模型不能被辨识,说明存在可产生简化式分布的多个参数解。在这种情况下,模型参数是非常任意的,而这对于科学推论来讲是有害的。特别是,这意味着参数的一致估计被排除在外(如 Gabrielsen,1978)。

等价这个概念与此相关,它是关于如何在经验上区分不同统计模型的概念。如果两个模型仅仅是参数的再表示,并具有相同的简化式分布,那么它们就是等价的,并且在任何数据上都是一致且相等的。而当模型表示不同的,甚至实质上相互矛盾的数据生成机制时,那么等价就是有害的。但从另一方面来说,我们有时可利用等价来简化估计问题,在 5.3.2 节将会予以说明。

在 5.2 节,我们先介绍一些关于辨识的有用的定义。然后,通过研究那些被假设为可辨识的简化式参数的映射属性及基本参数来对辨识进行分析性的评估。这个方法将会用多个模型来演示。最后,我们考察经验辨识问题。

在 5.3 节中,我们首先介绍等价的定义,并讨论等价的分析方法。这些方法包括分析在那些产生相同的简化式分布的参数化方案中是否存在一对一的转换关系。在用一些例子说明这个分析方法以后,我们还将考察经验等价问题。

我们所使用的方法将不那么形式化,我们关注于辨识与等价的分析在实践中如何处理,关于进一步的内容我们会给出参考文献。对于计算机求解,我们将特别考虑那些可直接用软件实现的方法,如 Mathematica (Wolfram,2003) 或 Maple (Maple 9 学习指南,2003)。

尽管读者们应该领会辨识与等价的重要性,但对于本章某些特别技术性的部分,可跳过不读。

## 5.2 辨　识

### 5.2.1 定　义

下述定义是有用的：

• 对于观测到的随机变量，当两个参数点 $\boldsymbol{\vartheta}^1$，$\boldsymbol{\vartheta}^2$ 都包含相同的简化式分布时，这两个参数点被称为**观测等价**；$g(\mathbf{y}|\mathbf{X};\boldsymbol{\vartheta}^1)=g(\mathbf{y}|\mathbf{X};\boldsymbol{\vartheta}^2)$。

• 对于任意参数点 $\boldsymbol{\vartheta}^1\in\boldsymbol{A}$，当不存在其他观测等价的参数点 $\boldsymbol{\vartheta}^2\in\boldsymbol{A}$ 时，参数向量 $\boldsymbol{\vartheta}\in\boldsymbol{A}$为全局可辨识。

• 当存在开邻域 $\boldsymbol{\vartheta}^0$，且不包含任何与 $\boldsymbol{\vartheta}^0$ 观测等价的其他 $\boldsymbol{\vartheta}$ 时，参数点 $\boldsymbol{\vartheta}^0\in\boldsymbol{A}$ 为局部可辨识。

应该指出的是，在参数空间 $\boldsymbol{A}$ 中任意位置的局部辨识是全局辨识的必要非充分条件，请见贝西格等人（Bechger et al.，2001：362）的例子。另外，本书所分析的模型一般指**非线性矩结构**。由此可见，在 $\boldsymbol{A}$ 中某个点的局部辨识并不意味着在 $\boldsymbol{A}$ 中任意点的局部辨识，并且这些被找到的参数点通常也不是可局部辨识的（请见 5.2.4 节的例子）。因此，本书所讨论的很多模型很可能是不可以全局辨识的，我们必须依赖于局部辨识这个较弱的概念（另见 McDonald，1982）。

### 5.2.2 局部辨识的解析性分析方法

回顾一下 3.3.2 节中的这个单维因子模型

$$y_{ij}=\beta_i+\lambda_i\eta_j+\epsilon_{ij},\quad \eta_j\sim \mathrm{N}(\gamma,\psi),\epsilon_{ij}\sim \mathrm{N}(0,\theta_{ii})$$

对因子进行线性转换，$f_j=a\eta_j+c$，该模型可重写为

$$\begin{aligned} y_{ij} &= (\beta_i-\lambda_i c/a)+(\lambda_i/a)f_j+\epsilon_{ij},\quad f_j\sim \mathrm{N}(a\gamma+c,a^2\psi),\quad \epsilon_{ij}\sim \mathrm{N}(0,\theta_{ii}) \\ &= \beta_i^*+\lambda_i^*f_j+\epsilon_{ij},\quad f_j\sim \mathrm{N}(\gamma^*,\psi^*),\quad \epsilon_{ij}\sim \mathrm{N}(0,\theta_{ii}) \end{aligned}$$

其中

$$\beta_i^*=\beta_i-\lambda_i c/a,\lambda_i^*=\lambda_i/a,\gamma^*=a\gamma+c,\psi^*=a^2\psi$$

因此，不同的参数点能产生相同的简化式分布，该模型无法辨识。为了辨识该模型，我们可以将因子的均值和方差分别约束为 1 和 0。尽管证明上述模型无法辨识相当容易，然而为了证明我们的参数约束使得模型可辨识却并不是那么直截了当。事实上，当该模型的题器少于 3 个时，很显然就无法辨识。

在适当的假设下，在某个参数点局部辨识的一个必要充分条件为（理论上的）信息矩阵在该点是非奇异的（如 Rothenberg，1971）。原则上说，这个条件可用于分析辨识，但这个方法在实践上通常不可行，因为在复杂模型中信息矩阵通常难以处理。

在这个特殊的例子中，像一般情形那样存在简化式参数（完整表征简化式分布），那么标准的辨识方法就关注于基本参数与简化式参数之间的映射上面（见 4.10 节）。该方法对于局部辨识可得到必要充分条件，它似乎由瓦尔德首先提出（Wald，1950）；请见费舍尔的综述（Fisher，1966）。杜帕科瓦和沃尔德（Dupacǒvá and Wold，1982）将这个思路应用到常规的带潜变量的结构方程模型中。

在本章中,我们将映射方法从普通的响应扩展到产生自跨越不同阈值的正态潜响应的二项响应或定序响应模型上,如带正态概率链接函数的模型。这是可行的,因为在这个例子中,简化式参数存在并且完整表征简化式分布。对于那些不存在能够完整表征简化式分布的简化式参数模型来说,考虑一下基本参数与简化式参数的一阶矩和二阶矩之间的映射可能会有所帮助,其背后的思路是依赖于高阶矩的辨识很可能非常脆弱。例如,对于二项响应或定序响应,如果与其相对照的正态概率模型无法辨识的话,则对数概率模型的辨识就很可能也非常脆弱(如 Rabe-Hesketh and Skrondal,2001)。

通过下述映射,基本参数 $\boldsymbol{\vartheta}\in \boldsymbol{A}$ 可得到简化式参数 $\mathbf{m}\in \boldsymbol{A}'$,即

$$m_s = h_s(\boldsymbol{\vartheta}), \quad 1 \leqslant s \leqslant S$$

其中,$h_s(\cdot)$,$1\leqslant s\leqslant S$ 为连续可微已知函数。

观测变量的概率分布仅仅通过该 $S$ 维的简化式参数向量 $\mathbf{m}$ 而依赖于基本参数 $\boldsymbol{\vartheta}\in \boldsymbol{A}$。对于所有的 $\boldsymbol{\vartheta}\in \boldsymbol{A}$,则

$$g(\mathbf{y}\mid \mathbf{X};\boldsymbol{\vartheta}) = g^*(\mathbf{y}\mid \mathbf{X};h_1(\boldsymbol{\vartheta}),\cdots,h_S(\boldsymbol{\vartheta})) = g^*(\mathbf{y}\mid \mathbf{X};\mathbf{m})$$

其中,$g^*$ 为简化式参数的分布。从而,$\boldsymbol{\vartheta}$ 的辨识可通过从 $\boldsymbol{\vartheta}$ 到 $\mathbf{m}$ 的映射特性来探究。

考虑一个特殊的基本参数向量 $\boldsymbol{\vartheta}^0$,它可得到简化式参数向量 $\mathbf{m}^0$,即

$$m_s^0 = h_s(\boldsymbol{\vartheta}^0), \quad 1 \leqslant s \leqslant S$$

那么,当且仅当 $\boldsymbol{\vartheta}^0$ 为下述方程的唯一解时,$\boldsymbol{\vartheta}^0$ 可辨识,即

$$m_s^0 = h_s(\boldsymbol{\vartheta}), \quad 1 \leqslant s \leqslant S \tag{5.1}$$

因此,$\boldsymbol{\vartheta}^0$ 的辨识完全依赖于映射 $h_s(\cdot)$ 的属性。那么,辨识的问题就成为求方程组的唯一解的问题,从而我们可使用微积分的常规结果。

很明显,关于辨识的必要非充分条件为简化式参数向量的元素个数小于或等于未知参数个数,即 $v\leqslant S$。为了推导出更有力的辨识结果,定义下面这个映射的雅可比行列式

$$J(\boldsymbol{\vartheta}) = \left[\frac{\partial h_s}{\partial \vartheta_j}, \quad 1 \leqslant s \leqslant S, 1 \leqslant j \leqslant v\right]$$

当 $\boldsymbol{\vartheta}^0$ 存在一个开邻域且雅可比行列式为常数秩时,参数向量 $\boldsymbol{\vartheta}^0$ 为**正则点**(regular point)。如果除了 $\boldsymbol{\vartheta}^0\in \boldsymbol{A}$ 之外,我们对 $\boldsymbol{\vartheta}^0$ 一无所知,那么可假设它为正则点,因为 $\boldsymbol{A}$ 的几乎所有的点都为正则点。然而,在 5.2.4 节中,我们将会碰到 $\boldsymbol{\vartheta}^0$ 为非正则点的情形。

如果 $\boldsymbol{\vartheta}^0$ 为正则点,当且仅当雅可比行列式的秩等于基本参数个数 $v$,方程组(5.1)有唯一解 $\boldsymbol{\vartheta}^0$。因此,本章关于辨识的分析就依赖于下面的引理:

**引理 1**:如果 $\boldsymbol{\vartheta}^0$ 为 $J(\boldsymbol{\vartheta})$ 的一个正则点,那么当且仅当 $\text{Rank}[J(\boldsymbol{\vartheta}^0)]=v$ 时,$\boldsymbol{\vartheta}^0$ 为可局部辨识。

## 5.2.3 雅可比法在局部辨识的应用

**包含 4 个连续题器的单因子模型**

我们回到 4.10 节所介绍的带有 4 个连续题器的单因子模型,其参数向量 $\boldsymbol{\vartheta}$ 及简化式参数 $\mathbf{m}(\boldsymbol{\vartheta})$ 分别由式(4.36) 和式(4.37)给出。该 $10\times 9$ 的雅可比行列式就为

$$\mathbf{J}(\boldsymbol{\vartheta}) = \begin{bmatrix} 2\psi\lambda_1 & 0 & 0 & 0 & \lambda_1^2 & 1 & 0 & 0 & 0 \\ \psi\lambda_2 & \psi\lambda_1 & 0 & 0 & \lambda_1\lambda_2 & 0 & 0 & 0 & 0 \\ \psi\lambda_3 & 0 & \psi\lambda_1 & 0 & \lambda_1\lambda_3 & 0 & 0 & 0 & 0 \\ \psi\lambda_4 & 0 & 0 & \psi\lambda_1 & \lambda_1\lambda_4 & 0 & 0 & 0 & 0 \\ 0 & 2\psi\lambda_2 & 0 & 0 & \lambda_2^2 & 0 & 1 & 0 & 0 \\ 0 & \psi\lambda_3 & \psi\lambda_2 & 0 & \lambda_2\lambda_3 & 0 & 0 & 0 & 0 \\ 0 & \psi\lambda_4 & 0 & \psi\lambda_2 & \lambda_2\lambda_4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2\psi\lambda_3 & 0 & \lambda_3^2 & 0 & 0 & 1 & 0 \\ 0 & 0 & \psi\lambda_4 & \psi\lambda_3 & \lambda_3\lambda_4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2\psi\lambda_4 & \lambda_4^2 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{5.2}$$

其中,处于行 $r$ 列 $c$ 的元素为 $\partial m_r(\boldsymbol{\vartheta})/\partial\vartheta_c$。该雅可比行列式的秩,Rank[$\mathbf{J}(\boldsymbol{\vartheta})$]为 8,即比参数个数少 1,因此,该模型不可局部辨识。

设 $\mathbf{N}(\boldsymbol{\vartheta})$ 为满足方程 $\mathbf{J}(\boldsymbol{\vartheta})\mathbf{N}(\boldsymbol{\vartheta}) = \mathbf{0}$ 的向量。这样的向量被称为雅可比行列式 $\mathbf{J}(\boldsymbol{\vartheta})$ 的零空间基。在这个例子中,向量

$$\left[-\frac{\lambda_1}{2\psi}, -\frac{\lambda_2}{2\psi}, -\frac{\lambda_3}{2\psi}, -\frac{\lambda_4}{2\psi}, 1, 0, 0, 0, 0\right]'$$

表示该零空间的一个基。我们可知,在因子负荷 $\lambda_i$ 和因子方差 $\psi$ 中,线性依赖是非常明显的,但在误差项方差 $\theta_{ii}$ 上并不是如此。重要的是,由此可得,$\theta_{ii}$ 可局部辨识,且可被一致估计,尽管该模型作为整体来说并不能被辨识。这通常称为**部分辨识**(partial identification)。为了辨识因子方差及负荷,将因子负荷或者因子方差约束为一个常数看上去似乎是一个好办法。

首先考虑一下这个因子模型在"锚定"之后,即将其中一个因子负荷约束为任意非零常数,是否可辨识,通常设 $\lambda_1 = 1$。对于剩余的 8 个未知参数,我们附加一个星号表示。通过在(5.2)中设 $\lambda_1 = 1$,忽略第 1 列因为该列包含关于 $\lambda_1$ 的导数不再为模型参数,从而得到 $10 \times 8$ 的雅可比行列式。所得到的雅可比行列式为

$$\mathbf{J}(\boldsymbol{\vartheta}^*) = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ \psi^* & 0 & 0 & \lambda_2^* & 0 & 0 & 0 & 0 \\ 0 & \psi^* & 0 & \lambda_3^* & 0 & 0 & 0 & 0 \\ 0 & 0 & \psi^* & \lambda_4^* & 0 & 0 & 0 & 0 \\ 2\psi^*\lambda_2^* & 0 & 0 & (\lambda_2^*)^2 & 0 & 1 & 0 & 0 \\ \psi^*\lambda_3^* & \psi^*\lambda_2^* & 0 & \lambda_2^*\lambda_3^* & 0 & 0 & 0 & 0 \\ \psi^*\lambda_4^* & 0 & \psi^*\lambda_2^* & \lambda_2^*\lambda_4^* & 0 & 0 & 0 & 0 \\ 0 & 2\psi^*\lambda_3^* & 0 & (\lambda_3^*)^2 & 0 & 0 & 1 & 0 \\ 0 & \psi^*\lambda_4^* & \psi^*\lambda_3^* & \lambda_3^*\lambda_4^* & 0 & 0 & 0 & 0 \\ 0 & 0 & 2\psi^*\lambda_4^* & (\lambda_4^*)^2 & 0 & 0 & 0 & 1 \end{bmatrix}$$

其秩 Rank[$\mathbf{J}(\boldsymbol{\vartheta}^*)$] = 8,因此在锚定之后,如果 $\boldsymbol{\vartheta}^*$ 为正则点,则该模型可局部辨识。

接下来,我们研究在"**因子标准化**(factor standardization)"之后,即将因子方差约束为任意非零常数,通常设为 $\psi = 1$,该因子模型是否可辨识。为了与锚定法中的参数相

区别,我们在符号上加一个横杠表示。在(5.2)中设 $\psi=1$,并忽略第5列,雅可比矩阵就成为

$$
\mathbf{J}(\bar{\boldsymbol{\vartheta}}) = \begin{bmatrix}
2\bar{\lambda}_1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
\bar{\lambda}_2 & \bar{\lambda}_1 & 0 & 0 & 0 & 0 & 0 & 0 \\
\bar{\lambda}_3 & 0 & \bar{\lambda}_1 & 0 & 0 & 0 & 0 & 0 \\
\bar{\lambda}_4 & 0 & 0 & \bar{\lambda}_1 & 0 & 0 & 0 & 0 \\
0 & 2\bar{\lambda}_2 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & \bar{\lambda}_3 & \bar{\lambda}_2 & 0 & 0 & 0 & 0 & 0 \\
0 & \bar{\lambda}_4 & 0 & \bar{\lambda}_2 & 0 & 0 & 0 & 0 \\
0 & 0 & 2\bar{\lambda}_3 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & \bar{\lambda}_4 & \bar{\lambda}_3 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 2\bar{\lambda}_4 & 0 & 0 & 0 & 1
\end{bmatrix}
$$

其秩 $\text{Rank}[\mathbf{J}(\bar{\boldsymbol{\vartheta}})]=8$,因此如果 $\bar{\boldsymbol{\vartheta}}$ 为正则点,该模型可局部辨识。

**包含6个连续题器及误差相关的结构方程模型**

现在我们分析一个来自惠顿等人(Wheaton,1977)所讨论过的面板数据的结构方程模型,该模型包含3个潜变量。该模型为递归模型,其中一个潜变量作为解释变量,而另两个潜变量表示在两个波次中的响应变量。每个潜变量都由两个题器来测量,并且每个因子都将其中一个题器作为锚。该模型具有一个重要的特点,即对于同一个题器的重复测量来说,其误差相关,这就使得依靠纸笔手算的方法来对辨识进行探析变得非常复杂(另见 Jöreskog and Sörbom,1989:173-174)。该模型的路径图如图5.1所示,其中 $\epsilon_{ij}$ 用小圆圈表示。

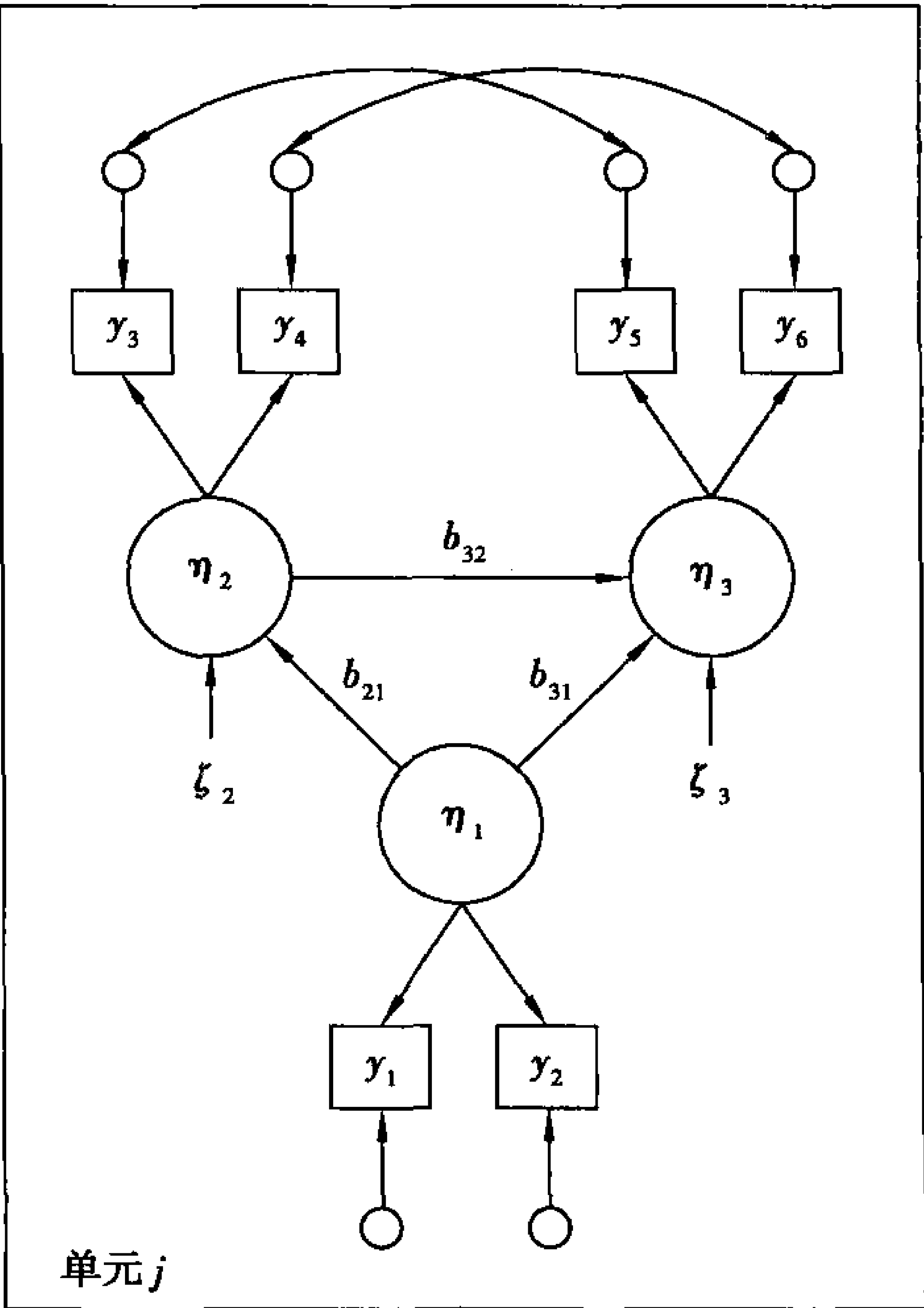



**图5.1 惠顿等人(Wheaton et al.,1977)的模型路径图**

该模型的测量部分为(对应于方程(3.31))

$$\begin{bmatrix} y_{1j} \\ y_{2j} \\ y_{3j} \\ y_{4j} \\ y_{5j} \\ y_{6j} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ \lambda_{21} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \lambda_{42} & 0 \\ 0 & 0 & 1 \\ 0 & 0 & \lambda_{63} \end{bmatrix} \begin{bmatrix} \eta_{1j} \\ \eta_{2j} \\ \eta_{3j} \end{bmatrix} + \begin{bmatrix} \epsilon_{1j} \\ \epsilon_{2j} \\ \epsilon_{3j} \\ \epsilon_{4j} \\ \epsilon_{5j} \\ \epsilon_{6j} \end{bmatrix}$$

其中

$$\begin{bmatrix} \epsilon_{1j} \\ \epsilon_{2j} \\ \epsilon_{3j} \\ \epsilon_{4j} \\ \epsilon_{5j} \\ \epsilon_{6j} \end{bmatrix} \sim \mathrm{N}_6\left( \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \theta_{11} & & & & & \\ 0 & \theta_{22} & & & & \\ 0 & 0 & \theta_{33} & & & \\ 0 & 0 & 0 & \theta_{44} & & \\ 0 & 0 & \theta_{53} & 0 & \theta_{55} & \\ 0 & 0 & 0 & \theta_{64} & 0 & \theta_{66} \end{bmatrix} \right)$$

结构部分(对应于方程(3.32))被设定为

$$\begin{bmatrix} \eta_{1j} \\ \eta_{2j} \\ \eta_{3j} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ b_{21} & 0 & 0 \\ b_{31} & b_{32} & 0 \end{bmatrix} \begin{bmatrix} \eta_{1j} \\ \eta_{2j} \\ \eta_{3j} \end{bmatrix} + \begin{bmatrix} \zeta_{1j} \\ \zeta_{2j} \\ \zeta_{3j} \end{bmatrix}$$

其中

$$\begin{bmatrix} \zeta_{1j} \\ \zeta_{2j} \\ \zeta_{3j} \end{bmatrix} \sim \mathrm{N}_3\left( \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \psi_{11} & & \\ 0 & \psi_{22} & \\ 0 & 0 & \psi_{33} \end{bmatrix} \right)$$

该模型具有 17 个参数,另外还暗含 21 个非冗余二阶矩。我们得到 $\mathrm{Rank}[\mathbf{J}(\boldsymbol{\vartheta})]=17$,因此当 $\boldsymbol{\vartheta}$ 为正则点时,该模型可局部辨识。

考虑到面板数据设计的因素,还可尝试让干扰项 $\zeta_2$ 与 $\zeta_3$ 相关,即 $\psi_{32}\neq 0$。然而,这将得到 $\mathrm{Rank}[\mathbf{J}(\boldsymbol{\vartheta})]=17$,而参数个数为 18 个,说明该模型无法局部辨识。

**包含 4 个二项题器的单因子模型**

请考虑下述包含 4 个连续潜响应变量的单因子模型:

$$\begin{bmatrix} y_{1j}^* \\ y_{2j}^* \\ y_{3j}^* \\ y_{4j}^* \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} + \begin{bmatrix} 1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \end{bmatrix} \eta_j + \begin{bmatrix} \epsilon_{1j} \\ \epsilon_{2j} \\ \epsilon_{3j} \\ \epsilon_{4j} \end{bmatrix}$$

其中,$\eta_j \sim \mathrm{N}(0,\psi)$,$\epsilon_{ij} \sim \mathrm{N}(0,\theta_{ii})$,以及 $\mathrm{Cov}(\epsilon_{ij},\epsilon_{i'j})=0$。请注意,我们已通过约束 $\lambda_1=1$ 将因子"锚定"了。

该模型包含 12 个未知参数

$$\boldsymbol{\vartheta} = [\beta_1,\beta_2,\beta_3,\beta_4,\lambda_2,\lambda_3,\lambda_4,\psi,\theta_{11},\theta_{22},\theta_{33},\theta_{44}]'$$

并且,潜响应变量的协方差矩阵变成

$$\boldsymbol{\Omega}(\boldsymbol{\vartheta}) = \begin{bmatrix} \psi + \theta_{11} & & & \\ \lambda_2\psi & \lambda_2^2\psi + \theta_{22} & & \\ \lambda_3\psi & \lambda_3\psi\lambda_2 & \lambda_3^2\psi + \theta_{33} & \\ \lambda_4\psi & \lambda_4\psi\lambda_2 & \lambda_4\psi\lambda_3 & \lambda_4^2\psi + \theta_{44} \end{bmatrix}$$

现在,这个模型与包含 4 个连续题器且截距为 $\beta_i$ 的单因子模型相同,但根本的区别在于它的潜响应变量未被观测到。另外,该潜响应变量与通过阈值函数所观测到的二项响应相关,即

$$y_{ij} = \begin{cases} 1 & 当 y_{ij}^* > 0 \\ 0 & 其他 \end{cases}$$

边缘概率 $\Pr(y_{ij}=1)$ 就成为

$$\Pr(y_{ij} = 1) = \frac{1}{\sqrt{\lambda_i^2\psi + \theta_{ii}}}\int_0^\infty \phi\left(\frac{\beta_i + \xi_{ij}}{\sqrt{\lambda_i^2\psi + \theta_{ii}}}\right)\mathrm{d}\xi_{ij} = \Phi\left(\frac{\beta_i}{\lambda_i^2\psi + \theta_{ii}}\right)$$

其中,$\phi$ 为标准正态密度,$\xi_{ij} = \lambda_i\eta_j + \epsilon_{ij}$,且 $\Phi$ 为累积标准正态分布。题器 $i$ 的潜响应 $y_{ij}^*$ 的均值 $\mu_i$ 为

$$\mu_i = \frac{\beta_i}{\sqrt{\lambda_i^2\psi + \theta_{ii}}}$$

并且根据边缘概率可辨识。联合响应概率 $\mathrm{p}_{stuv} \equiv \Pr(y_1 = s, y_2 = t, y_3 = u, y_4 = v)$ 可表示为

$$\mathrm{p}_{stuv} = \frac{1}{\sqrt{(\psi + \theta_{11})(\lambda_2^2\psi + \theta_{22})(\lambda_3^2\psi + \theta_{33})(\lambda_4^2\psi + \theta_{44})}}\int_{-\tau_{s-1}}^{-\tau_s}\int_{-\tau_{t-1}}^{-\tau_t}\int_{-\tau_{u-1}}^{-\tau_u}\int_{-\tau_{v-1}}^{-\tau_v} \times$$

$$\varphi\left(\frac{\beta_1 + \xi_1}{\sqrt{\psi + \theta_{11}}}, \frac{\beta_2 + \xi_2}{\sqrt{\lambda_2^2\psi + \theta_{22}}}, \frac{\beta_3 + \xi_3}{\sqrt{\lambda_3^2\psi + \theta_{33}}}, \frac{\beta_4 + \xi_4}{\sqrt{\lambda_4^2\psi + \theta_{44}}}; \mathbf{R}\right)\mathrm{d}\xi_1\mathrm{d}\xi_2\mathrm{d}\xi_3\mathrm{d}\xi_4$$

其中,$\varphi(.\,,.\,,.\,,.\,;\mathbf{R})$ 为四维的标准正态密度,其四项相关矩阵为

$$\mathbf{R}(\boldsymbol{\vartheta}) = \mathrm{diag}(\boldsymbol{\Omega}(\boldsymbol{\vartheta}))^{-1/2}\boldsymbol{\Omega}(\boldsymbol{\vartheta})\mathrm{diag}(\boldsymbol{\Omega}(\boldsymbol{\vartheta}))^{-1/2}$$

$$= \begin{bmatrix} 1 & & & \\ \frac{\lambda_2\psi}{\sqrt{\lambda_2^2\psi + \theta_{22}}\sqrt{\psi + \theta_{11}}} & 1 & & \\ \frac{\lambda_3\psi}{\sqrt{\lambda_3^2\psi + \theta_{33}}\sqrt{\psi + \theta_{11}}} & \frac{\lambda_3\psi\lambda_2}{\sqrt{\lambda_3^2\psi + \theta_{33}}\sqrt{\lambda_2^2\psi + \theta_{22}}} & 1 & \\ \frac{\lambda_4\psi}{\sqrt{\lambda_4^2\psi + \theta_{44}}\sqrt{\psi + \theta_{11}}} & \frac{\lambda_4\psi\lambda_2}{\sqrt{\lambda_4^2\psi + \theta_{44}}\sqrt{\lambda_2^2\psi + \theta_{22}}} & \frac{\lambda_4\psi\lambda_3}{\sqrt{\lambda_4^2\psi + \theta_{44}}\sqrt{\lambda_3^2\psi + \theta_{33}}} & 1 \end{bmatrix}$$

该四项相关矩阵可被辨识。

总而言之,共有 10 个简化式参数,4 个均值以及 6 个非冗余四项相关,可用下式表示

$$\mathbf{m}(\boldsymbol{\vartheta}) = \left[\frac{\beta_1}{\sqrt{\psi + \theta_{11}}}, \frac{\beta_2}{\sqrt{\lambda_2^2\psi + \theta_{22}}}, \frac{\beta_3}{\sqrt{\lambda_3^2\psi + \theta_{33}}}, \frac{\beta_4}{\sqrt{\lambda_4^2\psi + \theta_{44}}},\right.$$

$$\frac{\lambda_2\psi}{\sqrt{\lambda_2^2\psi + \theta_{22}}\sqrt{\psi + \theta_{11}}}, \frac{\lambda_3\psi}{\sqrt{\lambda_3^2\psi + \theta_{33}}\sqrt{\psi + \theta_{11}}}, \frac{\lambda_4\psi}{\sqrt{\lambda_4^2\psi + \theta_{44}}\sqrt{\psi + \theta_{11}}},$$

$$\left.\frac{\lambda_3\psi\lambda_2}{\sqrt{\lambda_3^2\psi + \theta_{33}}\sqrt{\lambda_2^2\psi + \theta_{22}}}, \frac{\lambda_4\psi\lambda_2}{\sqrt{\lambda_4^2\psi + \theta_{44}}\sqrt{\lambda_2^2\psi + \theta_{22}}}, \frac{\lambda_4\psi\lambda_3}{\sqrt{\lambda_4^2\psi + \theta_{44}}\sqrt{\lambda_3^2\psi + \theta_{33}}}\right]$$

在这个例子中,为了得到雅可比行列式的秩,我们无须更进一步讨论,因为未知参数个数(12 个)多于已辨识的简化式参数个数(10 个)。因此,很显然,该模型不可辨识。

现在,考虑一下将误差的方差$\epsilon_{ij}$约束为 1,即$\theta_{11}=\theta_{22}=\theta_{33}=\theta_{44}=1$。共有 8 个未知参数

$$\boldsymbol{\vartheta} = \left[\beta_1,\beta_2,\beta_3,\beta_4,\lambda_2,\lambda_3,\lambda_4,\psi\right]'$$

那么,这 10 个已辨识的简化式参数则为

$$\mathbf{m}(\boldsymbol{\vartheta}) = \left[\frac{\beta_1}{\sqrt{\psi+1}},\frac{\beta_2}{\sqrt{\lambda_2^2\psi+1}},\frac{\beta_3}{\sqrt{\lambda_3^2\psi+1}},\frac{\beta_4}{\sqrt{\lambda_4^2\psi+1}},\right.$$
$$\frac{\lambda_2\psi}{\sqrt{\lambda_2^2\psi+1}\sqrt{\psi+1}},\frac{\lambda_3\psi}{\sqrt{\lambda_3^2\psi+1}\sqrt{\psi+1}},\frac{\lambda_4\psi}{\sqrt{\lambda_4^2\psi+1}\sqrt{\psi+1}},$$
$$\frac{\lambda_3\psi\lambda_2}{\sqrt{\lambda_3^2\psi+1}\sqrt{\lambda_2^2\psi+1}},\frac{\lambda_4\psi\lambda_2}{\sqrt{\lambda_4^2\psi+1}\sqrt{\lambda_2^2\psi+1}},$$
$$\left.\frac{\lambda_4\psi\lambda_3}{\sqrt{\lambda_4^2\psi+1}\sqrt{\lambda_3^2\psi+1}}\right] \tag{5.3}$$

在这个例子中,这个 10×8 的雅可比行列式$\mathbf{J}(\boldsymbol{\vartheta})$变得过于复杂甚至太庞大,以致无法展示出来。但是,要点在于 Rank$[\mathbf{J}(\boldsymbol{\vartheta})]$为 8,因此,当$\boldsymbol{\vartheta}$为正则点时,该模型可局部辨识。

**库尔与阿奎斯提的包含 4 个二项响应的模型**

库尔与阿奎斯提提出了一个多元二项对数概率-正态模型(BLN)(Coull and Agresti, 2000)。在他们的第一个例子中,他们设定了一个简单模型,在 4 个波次中均包含一个单独的截距且无其他协变量。在此,我们将分析他们的正态概率-正态(probit-normal)模型,该模型在拉贝-赫斯基思和斯科隆多一书中被标记为 BPN 模型(Rabe-Hesketh and Skrondal,2001):

$$\begin{bmatrix} y_{1j}^* \\ y_{2j}^* \\ y_{3j}^* \\ y_{4j}^* \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} + \begin{bmatrix} \eta_{1j} \\ \eta_{2j} \\ \eta_{3j} \\ \eta_{4j} \end{bmatrix} + \begin{bmatrix} \epsilon_{1j} \\ \epsilon_{2j} \\ \epsilon_{3j} \\ \epsilon_{4j} \end{bmatrix}$$

其中

$$\begin{bmatrix} \eta_{1j} \\ \eta_{2j} \\ \eta_{3j} \\ \eta_{4j} \end{bmatrix} \sim \mathrm{N}_4\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix},\begin{bmatrix} \sigma^2 & & & \\ \rho_1\sigma^2 & \sigma^2 & & \\ \rho_1\sigma^2 & \rho_1\sigma^2 & \sigma^2 & \\ \rho_2\sigma^2 & \rho_2\sigma^2 & \rho_2\sigma^2 & \sigma^2 \end{bmatrix}\right)$$

$\epsilon_{ij}\sim \mathrm{N}(0,1)$且$\mathrm{Cov}(\epsilon_{ij},\epsilon_{i'j})=0$。潜响应变量通过阈值函数与观测到的二项响应相关联

$$y_{ij} = \begin{cases} 1 & \text{当 } y_{ij}^* > 0 \\ 0 & \text{其他} \end{cases}$$

由于潜响应误差的方差不可辨识,正如前面所示的二项题器单因子模型,该模型不可辨识。

拉贝-赫斯基思和斯科隆多(Rabe-Hesketh and Skrondal,2001)提出将$\sigma^2$约束为一个正数来确保辨识。事实上,$\sigma^2$无法被约束为任何正数(Rabe-Hesketh and Skrondal,2001:

1258);但在此为了简单起见,我们约束 $\sigma^2=1$。那么则有6个未知参数

$$\boldsymbol{\vartheta} = [\beta_1,\beta_2,\beta_3,\beta_4,\rho_1,\rho_2]'$$

即意味着下述四项相关矩阵

$$\mathbf{R}(\boldsymbol{\vartheta}) = \begin{bmatrix} 1 & & & \\ \frac{\rho_1}{2} & 1 & & \\ \frac{\rho_1}{2} & \frac{\rho_1}{2} & 1 & \\ \frac{\rho_2}{2} & \frac{\rho_2}{2} & \frac{\rho_2}{2} & 1 \end{bmatrix}$$

其中,非对角线的元素可辨识。根据边缘概率

$$\Pr(y_{ij}=1) = \Phi\left(\frac{\beta_i}{\sqrt{2}}\right)$$

我们可辨识

$$\mu_i = \frac{\beta_i}{\sqrt{2}}$$

这些已被辨识的简化式参数可替代为

$$\mathbf{m}(\boldsymbol{\vartheta}) = \left[\frac{\beta_1}{\sqrt{2}},\frac{\beta_2}{\sqrt{2}},\frac{\beta_3}{\sqrt{2}},\frac{\beta_4}{\sqrt{2}},\frac{\rho_1}{2},\frac{\rho_1}{2},\frac{\rho_2}{2},\frac{\rho_1}{2},\frac{\rho_2}{2},\frac{\rho_2}{2}\right] \tag{5.4}$$

从而 $10\times6$ 的雅可比行列式则为

$$\mathbf{J}(\boldsymbol{\vartheta}) = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} \end{bmatrix}$$

其中 $\text{Rank}[\mathbf{J}(\boldsymbol{\vartheta})]=6$,因此该模型可局部辨识。另请注意,$\boldsymbol{\vartheta}$ 并未包含在该雅可比矩阵中,因此,该模型在参数空间上可局部辨识。

**包含3个定序题器的单因子模型**

请考虑下面这个带有3个连续潜响应变量的单因子模型:

$$\begin{bmatrix} y_{1j}^* \\ y_{2j}^* \\ y_{3j}^* \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix} \eta_j + \begin{bmatrix} \epsilon_{1j} \\ \epsilon_{2j} \\ \epsilon_{3j} \end{bmatrix}$$

其中 $\eta_j \sim N(0,1)$，$\epsilon_{ij} \sim N(0,\theta_{ii})$，以及 $\mathrm{Cov}(\epsilon_{ij},\epsilon_{i'j})=0$。该潜响应变量通过阈值函数与观测到的三项响应(3 个类别)相关联，且在所有类别上为恒定阈值，即

$$y_{ij} = \begin{cases} 0 & 当 \quad y_{ij}^* \leqslant 0 \\ 1 & 当 \quad 0 < y_{ij}^* \leqslant \kappa_2 \\ 2 & 当 \quad \kappa_2 < y_{ij}^* \end{cases}$$

共有 10 个未知参数

$$\boldsymbol{\vartheta} = \left[\kappa_2, \beta_1, \beta_2, \beta_3, \lambda_1, \lambda_2, \lambda_3, \theta_{11}, \theta_{22}, \theta_{33}\right]'$$

其边缘概率为

$$\Pr(y_{ij}=0) = \Phi\left(\frac{\beta_i}{\sqrt{\lambda_i^2+\theta_{ii}}}\right) \tag{5.5}$$

$$\Pr(y_{ij}=1) = \Phi\left(\frac{\beta_i-\kappa_2}{\sqrt{\lambda_i^2+\theta_{ii}}}\right) - \Phi\left(\frac{\beta_i}{\sqrt{\lambda_i^2+\theta_{ii}}}\right)$$

且

$$\Pr(y_{ij}=2) = 1-\Phi\left(\frac{\beta_i-\kappa_2}{\sqrt{\lambda_i^2+\theta_{ii}}}\right) \tag{5.6}$$

根据式(5.5)，当 $i=1,2,3$ 时，可辨识其均值为

$$\mu_i = \frac{\beta_i}{\sqrt{\lambda_i^2+\theta_{ii}}}$$

根据式(5.6)，其阈值为

$$\tau_i = \frac{\kappa_2}{\sqrt{\lambda_i^2+\theta_{ii}}}$$

可被辨识。

多项相关矩阵则为

$$\mathbf{R}(\boldsymbol{\vartheta}) = \begin{bmatrix} 1 & & \\ \dfrac{\lambda_2\lambda_1}{\sqrt{\lambda_2^2+\theta_{22}}\sqrt{\lambda_1^2+\theta_{11}}} & 1 & \\ \dfrac{\lambda_3\lambda_1}{\sqrt{\lambda_3^2+\theta_{33}}\sqrt{\lambda_1^2+\theta_{11}}} & \dfrac{\lambda_3\lambda_2}{\sqrt{\lambda_3^2+\theta_{33}}\sqrt{\lambda_2^2+\theta_{22}}} & 1 \end{bmatrix}$$

其中，非冗余多项相关可辨识。这 9 个简化式参数(均值、阈值及多项相关)置入

$$\mathbf{m}(\boldsymbol{\vartheta}) = \left[\frac{\beta_1}{\sqrt{\lambda_1^2+\theta_{11}}}, \frac{\beta_2}{\sqrt{\lambda_2^2+\theta_{22}}}, \frac{\beta_3}{\sqrt{\lambda_3^2+\theta_{33}}}, \frac{\kappa_2}{\sqrt{\lambda_1^2+\theta_{11}}}, \frac{\kappa_2}{\sqrt{\lambda_2^2+\theta_{22}}}, \frac{\kappa_2}{\sqrt{\lambda_3^2+\theta_{33}}}, \right.$$
$$\left. \frac{\lambda_2\lambda_1}{\sqrt{\lambda_2^2+\theta_{22}}\sqrt{\lambda_1^2+\theta_{11}}}, \frac{\lambda_3\lambda_1}{\sqrt{\lambda_3^2+\theta_{33}}\sqrt{\lambda_1^2+\theta_{11}}}, \frac{\lambda_3\lambda_2}{\sqrt{\lambda_3^2+\theta_{33}}\sqrt{\lambda_2^2+\theta_{22}}}\right]$$

由于未知参数个数(10 个)多于已辨识的简化式参数个数(9 个)，很显然，该模型不可辨识。

为了使得模型可辨识，穆拉齐(Muraki,1990)针对上述模型进行了一个特殊处理，他将题器的误差方差约束为1($\theta_{11}=\theta_{22}=\theta_{33}=1$)。那么，现在就只剩下7个未知参数，即

$$\boldsymbol{\vartheta}=[\kappa_2,\beta_1,\beta_2,\beta_3,\lambda_1,\lambda_2,\lambda_3]'$$

这9个简化式参数置入

$$\mathbf{m}(\boldsymbol{\vartheta})=\left[\frac{\beta_1}{\sqrt{\lambda_1^2+1}},\frac{\beta_2}{\sqrt{\lambda_2^2+1}},\frac{\beta_3}{\sqrt{\lambda_3^2+1}},\frac{\kappa_2}{\sqrt{\lambda_1^2+1}},\frac{\kappa_2}{\sqrt{\lambda_2^2+1}},\frac{\kappa_2}{\sqrt{\lambda_3^2+1}},\right.$$
$$\left.\frac{\lambda_2\lambda_1}{\sqrt{\lambda_2^2+1}\sqrt{\lambda_1^2+1}},\frac{\lambda_3\lambda_1}{\sqrt{\lambda_3^2+1}\sqrt{\lambda_1^2+1}},\frac{\lambda_3\lambda_2}{\sqrt{\lambda_3^2+1}\sqrt{\lambda_2^2+1}}\right]$$

该9×7雅可比行列式的秩为Rank[$\mathbf{J}(\boldsymbol{\vartheta})$]=7，因此，当$\boldsymbol{\vartheta}$为正则点时，该模型可局部辨识。

斯科隆多(Skrondal,1996)认为穆拉齐(Muraki,1990)所使用的这些约束过于严格。他证明为了局部辨识只需将其中一个误差方差约束为1就足够了。为了不降低普适性，我们对第一个题器的误差方差进行约束($\theta_{11}=1$)。现在就有9个未知参数

$$\boldsymbol{\vartheta}=[\kappa_2,\beta_1,\beta_2,\beta_3,\lambda_1,\lambda_2,\lambda_3,\theta_{22},\theta_{33}]'$$

这9个已辨识的简化式参数置入

$$\mathbf{m}(\boldsymbol{\vartheta})=\left[\frac{\beta_1}{\sqrt{\lambda_1^2+1}},\frac{\beta_2}{\sqrt{\lambda_2^2+\theta_{22}}},\frac{\beta_3}{\sqrt{\lambda_3^2+\theta_{33}}},\frac{\kappa_2}{\sqrt{\lambda_1^2+1}},\frac{\kappa_2}{\sqrt{\lambda_2^2+\theta_{22}}},\frac{\kappa_2}{\sqrt{\lambda_3^2+\theta_{33}}},\right.$$
$$\left.\frac{\lambda_2\lambda_1}{\sqrt{\lambda_2^2+\theta_{22}}\sqrt{\lambda_1^2+1}},\frac{\lambda_3\lambda_1}{\sqrt{\lambda_3^2+\theta_{33}}\sqrt{\lambda_1^2+1}},\frac{\lambda_3\lambda_2}{\sqrt{\lambda_3^2+\theta_{33}}\sqrt{\lambda_2^2+\theta_{22}}}\right]$$

所得到的9×9雅可比行列式过于庞大，我们不将它显示在此，但其秩Rank[$\mathbf{J}(\boldsymbol{\vartheta})$]=9，因此，该模型正如斯科隆多(Skrondal,1996)所证明的，当$\boldsymbol{\vartheta}$为正则点时可局部辨识。在10.4节中，我们将采用这个参数化方法来研究美国人的生活满意度。

接下来，考虑一下这样一个模型，即如穆拉齐(Muraki,1990)那样将误差方差约束为1，但允许阈值在不同题器中间变动。该阈值函数则为

$$y_{ij}=\begin{cases}0 & 当\quad y_{ij}^*\leqslant 0\\ 1 & 当\quad 0<y_{ij}^*\leqslant\kappa_{i2}\\ 2 & 当\quad \kappa_{i2}<y_{ij}^*\end{cases}$$

其中，我们注意到阈值下标为$i$。共有9个未知参数

$$\boldsymbol{\vartheta}=[\kappa_{12},\kappa_{22},\kappa_{32},\beta_1,\beta_2,\beta_3,\lambda_1,\lambda_2,\lambda_3]'$$

这9个简化式参数处于

$$\mathbf{m}(\boldsymbol{\vartheta})=\left[\frac{\beta_1}{\sqrt{\lambda_1^2+1}},\frac{\beta_2}{\sqrt{\lambda_2^2+1}},\frac{\beta_3}{\sqrt{\lambda_3^2+1}},\frac{\kappa_{12}}{\sqrt{\lambda_1^2+1}},\frac{\kappa_{22}}{\sqrt{\lambda_2^2+1}},\frac{\kappa_{32}}{\sqrt{\lambda_3^2+1}},\right.$$
$$\left.\frac{\lambda_2\lambda_1}{\sqrt{\lambda_2^2+1}\sqrt{\lambda_1^2+1}},\frac{\lambda_3\lambda_1}{\sqrt{\lambda_3^2+1}\sqrt{\lambda_1^2+1}},\frac{\lambda_3\lambda_3}{\sqrt{\lambda_3^2+1}\sqrt{\lambda_2^2+1}}\right]$$

这个9×9的雅可比行列式的秩为Rank[$\mathbf{J}(\boldsymbol{\vartheta})$]=9，因此，当$\boldsymbol{\vartheta}$为正则点时，该模型可局部辨识。

## 5.2.4 正则点与局部辨识

**4个连续题器的双因子模型**

请考虑下述带有4个连续题器的双因子模型，且这两个因子都已“锚定”：

$$\begin{bmatrix} y_{1j} \\ y_{2j} \\ y_{3j} \\ y_{4j} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \lambda_{21} & 0 \\ 0 & 1 \\ 0 & \lambda_{42} \end{bmatrix} \begin{bmatrix} \eta_{1j} \\ \eta_{2j} \end{bmatrix} + \begin{bmatrix} \epsilon_{1j} \\ \epsilon_{2j} \\ \epsilon_{3j} \\ \epsilon_{4j} \end{bmatrix}$$

其中，$\begin{bmatrix} \eta_{1j} \\ \eta_{2j} \end{bmatrix} \sim \mathrm{N}_2\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \psi_{11} & \\ \psi_{21} & \psi_{22} \end{bmatrix}\right)$，$\epsilon_{ij} \sim \mathrm{N}(0, \theta_{ii})$，并且 $\mathrm{Cov}(\epsilon_{ij}, \epsilon_{i'j}) = 0$。该模型的路径图显示在图 5.2 的左半部分。

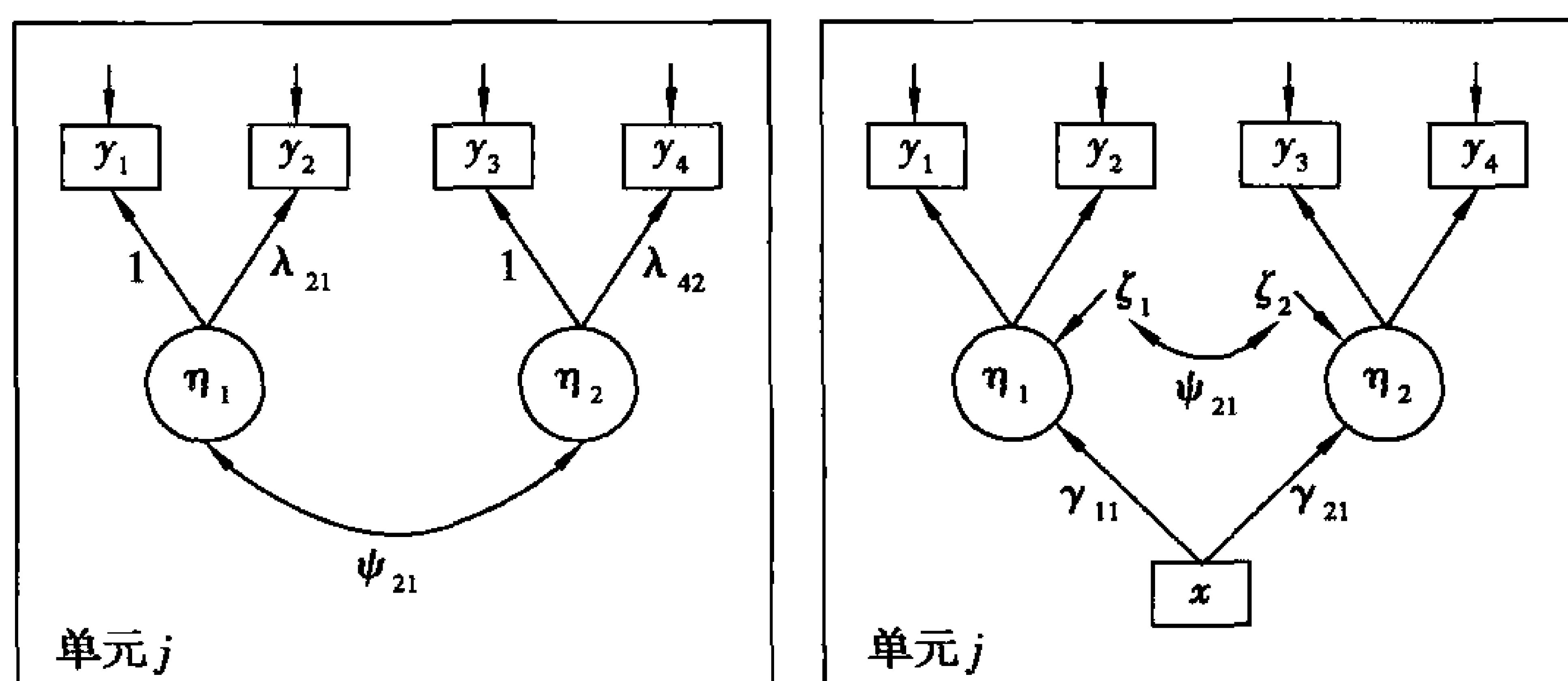



**图 5.2 包含及不包含外生变量的双因子模型**

在列向量 $\boldsymbol{\vartheta}$ 中共有 9 个未知参数

$$\boldsymbol{\vartheta} = [\lambda_{21}, \lambda_{42}, \psi_{11}, \psi_{21}, \psi_{22}, \theta_{11}, \theta_{22}, \theta_{33}, \theta_{44}]'$$

这意味着协方差矩阵为

$$\boldsymbol{\Omega}(\boldsymbol{\vartheta}) = \begin{bmatrix} \psi_{11} + \theta_{11} & & & \\ \lambda_{21}\psi_{11} & \lambda_{21}^2\psi + \theta_{22} & & \\ \psi_{21} & \psi_{21}\lambda_{21} & \psi_{22} + \theta_{33} & \\ \lambda_{42}\psi_{21} & \lambda_{42}\psi_{21}\lambda_{21} & \lambda_{42}\psi_{22} & \lambda_{42}^2\psi_{22} + \theta_{44} \end{bmatrix}$$

这 10 个简化式参数（非冗余方差和协方差）置入

$$\mathbf{m}(\boldsymbol{\vartheta}) = \big[\psi_{11} + \theta_{11}, \lambda_{21}\psi_{11}, \psi_{21}, \lambda_{42}\psi_{21}, \lambda_{21}^2\psi_{11} + \theta_{22}, \psi_{21}\lambda_{21}, \lambda_{42}\psi_{21}\lambda_{21},$$

$$\psi_{22} + \theta_{33}, \lambda_{42}\psi_{22}, \lambda_{42}^2\psi_{22} + \theta_{44}\big]$$

10 × 9 的雅可比行列式则为

$$\mathbf{J}(\boldsymbol{\vartheta}) = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ \psi_{11} & 0 & \lambda_{21} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \psi_{21} & 0 & \lambda_{42} & 0 & 0 & 0 & 0 & 0 \\ 2\lambda_{21}\psi_{11} & 0 & \lambda_{21}^2 & 0 & 0 & 0 & 1 & 0 & 0 \\ \psi_{21} & 0 & 0 & \lambda_{21} & 0 & 0 & 0 & 0 & 0 \\ \lambda_{42}\psi_{21} & \lambda_{21}\psi_{21} & 0 & \lambda_{21}\lambda_{42} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & \psi_{22} & 0 & 0 & \lambda_{42} & 0 & 0 & 0 & 0 \\ 0 & 2\lambda_{42}\psi_{22} & 0 & 0 & \lambda_{42}^2 & 0 & 0 & 0 & 1 \end{bmatrix}$$

其秩为 Rank[$\mathbf{J}(\boldsymbol{\vartheta})$] =9,因此,当 $\boldsymbol{\vartheta}$ 为正则点时,该模型可局部辨识。

现在让我们考虑一个特殊的情形,设因子之间互不相关,$\psi_{21}=0$。将该约束代入上述雅可比行列式,得到秩为7。因而,很明显,当 $\psi_{21}=0$ 时,$\vartheta^0$ 不是正则点,因为在此情形中,雅可比行列式在 $\boldsymbol{\vartheta}^0$ 的邻域上不是常数。并且,当 $\psi_{21}=0$ 时,该模型不可局部辨识,因为雅可比行列式的秩为7,而未知参数个数为8个($\lambda_{21},\lambda_{42},\psi_{11},\psi_{22},\theta_{11},\theta_{22},\theta_{33},\theta_{44}$)。因此,这是一个不太常见的情形,即释放(而非增加)一个参数约束才能使得模型局部辨识。该模型也非常好地说明,只有当参数点为**正则点**时,引理1才适用。

**4 个连续题器及外生变量的双因子模型**

现在我们对上述模型进行扩展,将因子对协变量 $x_j$ 回归

$$\begin{bmatrix}\eta_{1j}\\ \eta_{2j}\end{bmatrix}=\begin{bmatrix}\gamma_{11}\\ \gamma_{21}\end{bmatrix}x_j+\begin{bmatrix}\zeta_{1j}\\ \zeta_{2j}\end{bmatrix}$$

其中,$\begin{bmatrix}\zeta_{1j}\\ \zeta_{2j}\end{bmatrix}\sim N_2\left(\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix}\psi_{11} & \\ \psi_{21} & \psi_{22}\end{bmatrix}\right)$。该模型的路径图显示在图5.2的右半部分。

未知参数为

$$\boldsymbol{\vartheta}=[\gamma_{11},\gamma_{21},\lambda_{21},\lambda_{42},\psi_{11},\psi_{21},\psi_{22},\theta_{11},\theta_{22},\theta_{33},\theta_{44}]'$$

我们可得到回归或均值结构为

$$\begin{aligned}E(y_{1j}\mid x_j)&=\gamma_{11}x_j\\ E(y_{2j}\mid x_j)&=\lambda_{21}\gamma_{11}x_j\\ E(y_{3j}\mid x_j)&=\gamma_{21}x_j\\ E(y_{4j}\mid x_j)&=\lambda_{42}\gamma_{21}x_j\end{aligned}$$

那么,简化式参数向量就为

$$\mathbf{m}(\boldsymbol{\vartheta})=\Big[\gamma_{11},\lambda_{21}\gamma_{11},\gamma_{21},\lambda_{42}\gamma_{21},\psi_{11}+\theta_{11},\lambda_{21}\psi_{11},\psi_{21},\lambda_{42}\psi_{21},\lambda_{21}^2\psi_{11}+\theta_{22},$$
$$\psi_{21}\lambda_{21},\lambda_{42}\psi_{21}\lambda_{21},\psi_{22}+\theta_{33},\lambda_{42}\psi_{22},\lambda_{42}^2\psi_{22}+\theta_{44}\Big]$$

这个14×11的雅可比行列式则为

$$\mathbf{J}(\boldsymbol{\vartheta})=\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
\lambda_{21} & 0 & \gamma_{11} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & \lambda_{42} & 0 & \gamma_{21} & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0\\
0 & 0 & \psi_{11} & 0 & \lambda_{21} & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & \psi_{21} & 0 & \lambda_{42} & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 2\lambda_{21}\psi_{11} & 0 & \lambda_{21}^2 & 0 & 0 & 0 & 1 & 0 & 0\\
0 & 0 & \psi_{21} & 0 & 0 & \lambda_{21} & 0 & 0 & 0 & 0 & 0\\
0 & 0 & \lambda_{42}\psi_{21} & \lambda_{21}\psi_{21} & 0 & \lambda_{21}\lambda_{42} & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0\\
0 & 0 & 0 & \psi_{22} & 0 & 0 & \lambda_{42} & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 2\lambda_{42}\psi_{22} & 0 & 0 & \lambda_{42}^2 & 0 & 0 & 0 & 1
\end{bmatrix}$$

其秩 Rank[$\mathbf{J}(\boldsymbol{\vartheta})$] = 11，因此当 $\boldsymbol{\vartheta}$ 为正则点时，该模型可局部辨识。接下来请再考虑一个特殊情形，设 $\psi_{21}=0$；即这两个因子互不相关。将该约束代入上述雅可比行列式，其秩仍然为 11，因此，$\psi_{21}=0$ 并不意味着 $\boldsymbol{\vartheta}$ 为非正则点，与不包含协变量的情形相反。

## 5.2.5 经验辨识

解析性辨识通过未知的真实参数 $\boldsymbol{\vartheta}$ 进行。对于这个"理论"方法的一个有用的补充为基于所估计参数属性的"经验"辨识。由于经验辨识基于所估计的参数而非理论参数，它不如解析性方法那么严格，但是经验方法仍然具有一些优势：首先，经验分析基于所估计的信息矩阵，即最大似然估计的天然副产品。其次，经验辨识更加通用，因为它并不依赖于可全局辨识的简化式参数是否存在，而这些参数能够完整表征简化式分布。再次，有观点认为，经验辨识在参数估计时，能够对辨识进行评估。如，在 5.2.4 节中，如果 $\hat{\psi}_{21}\approx 0$，那么对于该双因子模型的推论将会存在问题。最后，经验辨识能够处理推论所基于的样本自身可能存在的问题。在线性回归中，预测变量之间的共线性就是一个经验辨识的例子。根据 Wiley (1973) 以及 McDonald and Krane (1977)，我们提出如下定义：

- 对于一个样本，如果最大似然解 $\hat{\boldsymbol{\vartheta}}$ 所估计的信息矩阵为非奇异时，那么模型可**经验辨识**。

请注意，这个条件仅仅是前面所提的基于理论信息矩阵条件（如 Rothenberg，1971）的一个经验副本。

矩阵接近奇异性程度的一种测量是**条件数**（condition number），其定义为最大的本征值与最小的本征值（eigenvalue）之比值的平方根。在操作中，如果条件数"过大"，超过某个阈值，我们将这个模型称为**经验欠缺辨识**（empirically underidentified）。如果一个模型经验欠缺辨识，参数估计的标准误以及组间相关值将会很大。例如，当在预测变量中间存在共线性时，以及在 5.2.4 节的例子中，如果双因子模型的 $\hat{\psi}_{21}\approx 0$，我们就可预计出现这个情形。

**二项对数概率-正态模型（BLN）**

我们现在来讨论库尔和阿奎斯提（Coull and Agresti, 2000）所讨论过的 BLN 模型，即上面所讨论的 BPN 模型的对数概率方案。拉贝-赫斯基思和斯科隆多（Rabe-Hesketh and Skrondal，2001）认为，根据一阶矩和二阶矩，BLN 模型无法辨识，从而很可能是经验欠缺辨识，因为高阶矩的信息很可能非常不充分。

如果不约束 $\sigma^2$，对模型进行估计，得到的条件数为 179.5，该值非常大（最小的本征值小于 0.004），该条件数说明观测到的信息矩阵接近于奇异矩阵。因此，BLN 模型似乎无法经验辨识。我们又估计了另外一个模型，将 $\sigma^2$ 约束为它的最大似然估计值 4.06，得到的条件数为 5.2。

将所估计的信息矩阵逆转，我们得到参数估计的协方差矩阵估计。从表 5.1 可知，当 $\sigma$ 被固定时，所估计的标准误大大降低。参数估计的相关系数显示在表 5.2 中。对于无约束模型，参数估计存在高度的组间相关，大多数相关系数接近于 ±1，最小的相关系数（绝对值）为 0.79，而在受约束模型中，最大的相关系数为 0.19。

在介绍了经验欠缺辨识之后，我们现在研究它是否由于高阶矩信息的不充分而造成的。在 $\sigma$ 的取值范围内，我们计算出其他参数来保留最大似然解所暗含的均值和相关。用这套不同的参数值表示的模型意味着相同的潜响应的一阶矩和二阶矩，但在高阶矩上不同。这些模型的偏差与 $\sigma$ 显示在图 5.3 中，其中 $\sigma$ 从 1.35 增加到 8，而该最小值与潜响应之间的相关一致。对于 $\sigma$ 的每个值，为了求得其他参数，在高阶矩中几乎没有任何变化的偏差被我们故意忽略掉了。这也就是潜响应的高阶矩中信息不充分的直接证据。

请注意,由于其他参数并不是通过最大似然法所估计的,图 5.3 中的曲线表示与 $\sigma$ 的剖面似然(见 8.3.5 节)相对应的偏差的上限。例如,将 $\sigma$ 固定为 8.2,估计模型所得到的偏差仅为 6.53。

**表 5.1 BLN 约束模型和无约束模型的参数估计、标准误以及偏差(每个维度 20 个积分点)**

| | 估计值 | 标准误差 | |
|---|---|---|---|
| | | 无约束模型 | 约束模型 |
| $\beta_1$ | -4.04 | 6.85 | 0.39 |
| $\beta_2$ | -4.42 | 7.34 | 0.41 |
| $\beta_3$ | -4.69 | 7.77 | 0.42 |
| $\beta_4$ | -4.56 | 7.57 | 0.42 |
| $\sigma$ | 4.06 | 7.99 | — |
| $\rho_1$ | 0.43 | 0.28 | 0.10 |
| $\rho_2$ | -0.25 | 0.20 | 0.12 |
| 偏差 | 6.28 | | |

资料来源:Rabe-Hesketh and Skrondal (2001)。

**表 5.2 BLN 模型参数估计相关系数矩阵;对角线上方:约束模型;对角线下方:无约束模型**

| | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_4$ | $\sigma$ | $\rho_1$ | $\rho_2$ |
|---|---|---|---|---|---|---|---|
| $\beta_1$ | 1 | 0.190 | 0.187 | -0.083 | — | 0.050 | -0.025 |
| $\beta_2$ | 0.997 | 1 | 0.185 | -0.080 | — | 0.062 | -0.033 |
| $\beta_3$ | 0.997 | 0.998 | 1 | -0.077 | — | 0.069 | -0.037 |
| $\beta_4$ | 0.997 | 0.997 | 0.997 | 1 | — | 0.014 | -0.045 |
| $\sigma$ | -0.998 | -0.998 | -0.999 | -0998 | 1 | — | — |
| $\rho_1$ | 0.941 | 0.942 | 0.942 | 0.941 | -0.942 | 1 | -0.106 |
| $\rho_2$ | -0.814 | -0.815 | -0.815 | -0.815 | 0.815 | -0.788 | 1 |

资料来源:Rabe-Hesketh and Skrondal (2001)。

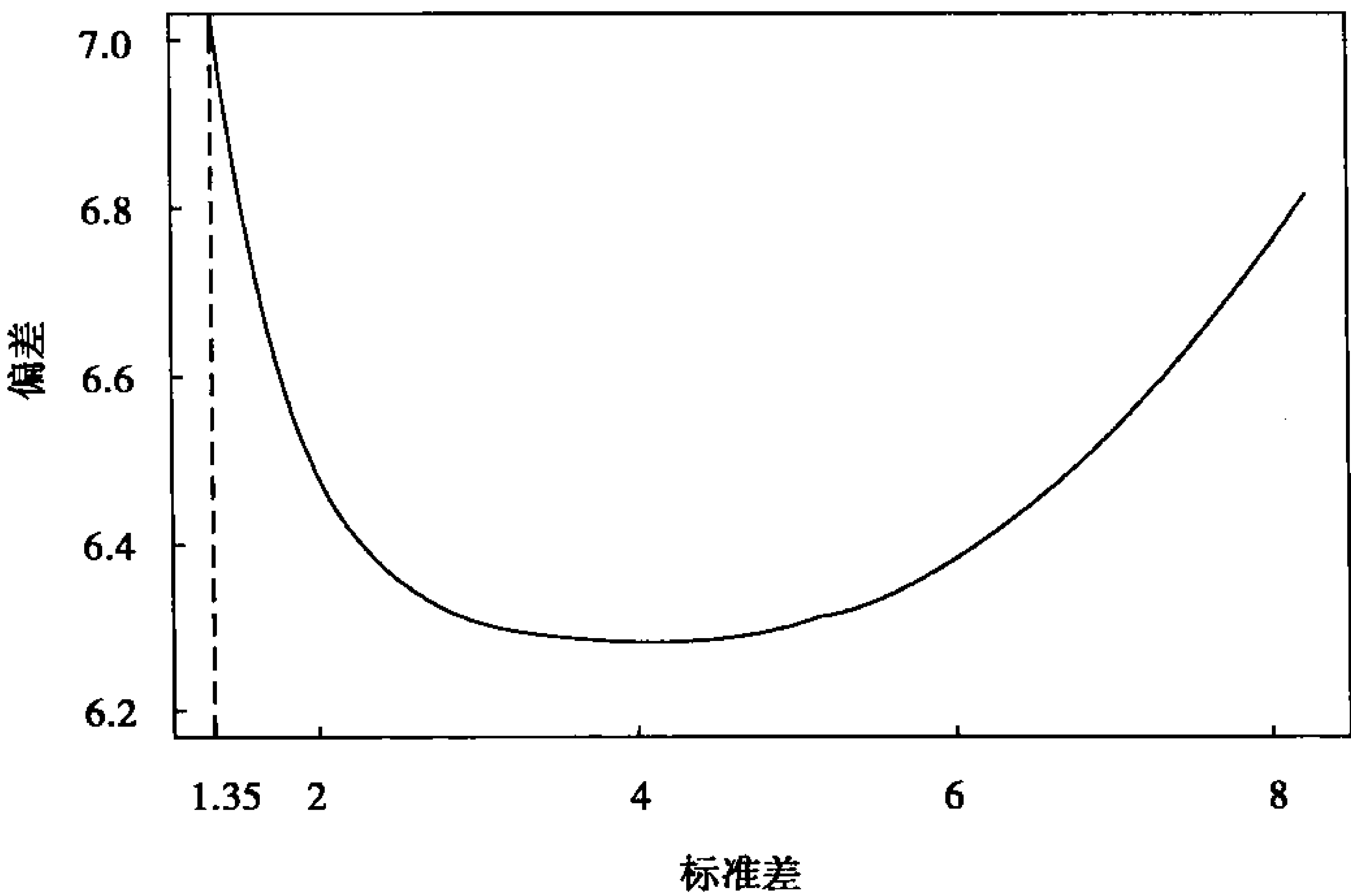



**图 5.3 标准差 $\sigma$ 不同取值下的偏差。计算其他参数以保留最大似然解所包含的一阶矩和二阶矩**

(资料来源:Rabe-Hesketh and Skrondal,2001)

## 5.3 等 价

### 5.3.1 定 义

• 两个统计模型 $M_1$ 与 $M_2$，有 $\boldsymbol{\vartheta}_A^1 \in \boldsymbol{A}$ 与 $\boldsymbol{\vartheta}_B^2 \in \boldsymbol{B}$，如果这两个模型通过参数优化，在 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 范围内存在关于 $\boldsymbol{\vartheta}_A$ 和 $\boldsymbol{\vartheta}_B$ 的一对一的转换，那么这两个模型就为全局等价。

正如在非线性矩结构中的辨识一样，全局等价的前景看起来也很黯淡，我们转向局部等价的概念：

• 分别具有可局部辨识的参数点 $\boldsymbol{\vartheta}_A^1 \in \boldsymbol{A}$ 与 $\boldsymbol{\vartheta}_B^2 \in \boldsymbol{B}$ 的两个统计模型 $M_1$ 与 $M_2$，如果它们通过参数优化后，在点的开邻域 $\boldsymbol{\vartheta}_A$ 和 $\boldsymbol{\vartheta}_B$ 之间存在一对一的转换，使得它们观测等价，则这两个统计模型 $M_1$ 与 $M_2$ 是局部等价。

### 5.3.2 等价的解析性分析

我们对于可用低阶矩完整表征的模型提出如下方法。请考虑两个基本参数向量为 $\boldsymbol{\vartheta}_A$ 与 $\boldsymbol{\vartheta}_B$ 的潜在等价模型，它们的简化式参数分别为 $m_s = h_s(\boldsymbol{\vartheta}_A)$ 和 $m_s^* = h_s^*(\boldsymbol{\vartheta}_B)$。如果这两个模型等价，由此可得，对于每个参数点 $\boldsymbol{\vartheta}_A^0$，存在参数点 $\boldsymbol{\vartheta}_B^0$，使得

$$h_s(\boldsymbol{\vartheta}_A^0) = h_s^*(\boldsymbol{\vartheta}_B^0)$$

那么我们就可用 $\boldsymbol{\vartheta}_B$ 来求解 $\boldsymbol{\vartheta}_A$ 以及用 $\boldsymbol{\vartheta}_A$ 来求解 $\boldsymbol{\vartheta}_B$，从而可以分析这两个模型的基本参数之间是否存在一对一的转换，以及在什么条件下存在。

对于来自同一个欠缺辨识的模型 $M_{12}$ 的两个子模型 $M_1$ 和 $M_2$，我们需要研究它们的局部等价，我们考察由鲁基本(Luijben,1991)所提出的方法。设 $\boldsymbol{\vartheta}^{12}$ 表示 $M_{12}$ 的基本参数 $v$ 的向量。定义模型 $M_0$，并且其约束为 $M_{12}$ 引致 $M_1$ 和 $M_2$ 的约束。另外，设 $\boldsymbol{\vartheta}_R^{12}$，假设它为 $M_{12}$ 的一个正则点，它为 $\boldsymbol{\vartheta}^{12}$ 的一个约束方案且此约束为引致 $M_0$ 的约束。在引致 $M_0$ 的约束条件下，$M_1$ 和 $M_2$ 的参数向量被假设为可局部辨识并且是正则点。定义雅可比行列式 $\mathbf{J}(\boldsymbol{\vartheta}^{12}) \mid \boldsymbol{\vartheta}^{12} = \boldsymbol{\vartheta}_R^{12}$ 为基本参数 $\boldsymbol{\vartheta}^{12}$ 在增加了 $\boldsymbol{\vartheta}_R^{12}$ 的约束后，映射到简化式参数向量 $\mathbf{m}^{12}$ 上的雅可比行列式 $\mathbf{J}(\boldsymbol{\vartheta}^{12})$。

那么在现在这个设置下，对等价的分析就依赖于下述引理：

**引理** 2：$M_1$ 和 $M_2$ 为**局部等价**当且仅当

$$\text{Rank}\left[\mathbf{J}(\boldsymbol{\vartheta}^{12}) \mid \boldsymbol{\vartheta}^{12} = \boldsymbol{\vartheta}_R^{12}\right] < v$$

正如在辨识中所采用的雅可比策略，这个研究等价的方法只适用于那些简化式参数能够完整表征简化式分布的模型。

然而，当这两个潜在的等价子模型所嵌套于其中的欠缺辨识的公共模型未知时，我们无法使用引理 2。贝克尔等(Bekker et al.,1994)指出，鲁基本的方法限制过多，他认为应当研究一般情形下的局部等价。但是，所推导出来的条件在操作中非常难以评估。

**包含 4 个连续题器的单因子模型**

让我们再次回到 4.10 节所介绍的带有 4 个连续题器的单因子模型。在 5.2.3 节，我们介绍了两个可局部辨识的模型，其中一个模型使用锚定法($\lambda_1 = 1$)，而另一个模型则使用了因子标准化法($\psi = 1$)。

我们使通过两种参数表达法所得到的简化式参数相等：

$$\mathbf{m}(\bar{\boldsymbol{\vartheta}}) = \left[\bar{\lambda}_1^2 + \bar{\theta}_{11}, \bar{\lambda}_2\bar{\lambda}_1, \bar{\lambda}_3\bar{\lambda}_1, \bar{\lambda}_4\bar{\lambda}_1, \bar{\lambda}_2^2 + \bar{\theta}_{22}, \bar{\lambda}_3\bar{\lambda}_2, \bar{\lambda}_4\bar{\lambda}_2, \bar{\lambda}_3^2 + \bar{\theta}_{33}, \bar{\lambda}_4\bar{\lambda}_3, \bar{\lambda}_4^2 + \bar{\theta}_{44}\right]$$

$$= \left[\psi^* + \theta_{11}^*, \lambda_2^*\psi^*, \lambda_3^*\psi^*, \lambda_4^*\psi^*, (\lambda_2^*)^2\psi^* + \theta_{22}^*, \lambda_3^*\psi^*\lambda_2^*, \lambda_4^*\psi^*\lambda_2^*, (\lambda_3^*)^2\psi^* + \theta_{33}^*, \lambda_4^*\psi^*\lambda_3^*, (\lambda_4^*)^2\psi^* + \theta_{44}^*\right]$$

$$= \mathbf{m}(\boldsymbol{\vartheta}^*)$$

对 $\boldsymbol{\vartheta}^*$ 求解,可得到唯一解

$$\left\{\theta_{11}^* = \bar{\theta}_{11}, \theta_{22}^* = \bar{\theta}_{22}, \theta_{33}^* = \bar{\theta}_{33}, \theta_{44}^* = \bar{\theta}_{44}, \lambda_2^* = \frac{\bar{\lambda}_2}{\bar{\lambda}_1}, \lambda_3^* = \frac{\bar{\lambda}_3}{\bar{\lambda}_1}, \lambda_2^* = \frac{\bar{\lambda}_4}{\bar{\lambda}_1}, \psi^* = \bar{\lambda}_1^2\right\}$$

当 $\bar{\boldsymbol{\lambda}}_1 \neq 0$,这个标准化了的模型显然能够在锚定模型的参数空间上生成参数。

对 $\bar{\boldsymbol{\vartheta}}$ 求解,我们可得到唯一解,并以因子负荷 $\bar{\lambda}_i$ 的符号方向为条件,即

$$\left\{\bar{\theta}_{11} = \theta_{11}^*, \bar{\theta}_{22} = \theta_{22}^*, \bar{\theta}_{33} = \theta_{33}^*, \bar{\theta}_{44} = \theta_{44}^*, \bar{\lambda}_1 = \pm\sqrt{\psi^*}, \bar{\lambda}_2 = \pm\lambda_2^*\sqrt{\psi^*}, \bar{\lambda}_3 = \pm\lambda_3^*\sqrt{\psi^*}, \bar{\lambda}_4 = \pm\lambda_4^*\sqrt{\psi^*}\right\}$$

如果 $\psi^* \geqslant 0$,那么该锚定模型似乎也能在标准化模型的参数空间上生成参数。

请注意,这两个模型都嵌套于无约束模型。因此,我们可以应用引理 2 来分析它们是否局部等价。将约束 $\lambda_1 = 1$(锚定)和 $\psi = 1$(因子标准化)代入式(5.2)中的未被辨识模型的雅可比行列式,得

$$\mathbf{J}(\boldsymbol{\vartheta}^{12}) \mid \boldsymbol{\vartheta}^{12} = \boldsymbol{\vartheta}_R^{12} = \begin{bmatrix} 2 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ \lambda_2 & 1 & 0 & 0 & \lambda_2 & 0 & 0 & 0 & 0 \\ \lambda_3 & 0 & 1 & 0 & \lambda_3 & 0 & 0 & 0 & 0 \\ \lambda_4 & 0 & 0 & 1 & \lambda_4 & 0 & 0 & 0 & 0 \\ 0 & 2\lambda_2 & 0 & 0 & \lambda_2^2 & 0 & 1 & 0 & 0 \\ 0 & \lambda_3 & \lambda_2 & 0 & \lambda_2\lambda_3 & 0 & 0 & 0 & 0 \\ 0 & \lambda_4 & 0 & \lambda_2 & \lambda_2\lambda_4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2\lambda_3 & 0 & \lambda_3^2 & 0 & 0 & 1 & 0 \\ 0 & 0 & \lambda_4 & \lambda_3 & \lambda_3\lambda_4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2\lambda_4 & \lambda_4^2 & 0 & 0 & 0 & 1 \end{bmatrix}$$

该雅可比行列式的秩 Rank $\left[\mathbf{J}(\boldsymbol{\vartheta}^{12}) \mid \boldsymbol{\vartheta}^{12} = \boldsymbol{\vartheta}_R^{12}\right]$ 为 8,比参数个数少 1,因此,在上面的假设下,这两个模型局部等价。

尽管在参数优化上等价,锚定法通常更可能被选用而非标准化法,因为它确保了因子恒等性(见 3.3.2 节)。

**BPN 模型及受约束二项单因子模型**

现在我们分析 5.2.3 节所示的已被辨识的 BPN 模型是否等价于在同一节所介绍的带有 4 个二项题器的受约束单因子模型。在这个例子中,我们未发现任何这两个模型所嵌套于其中的公共的欠缺辨识的模型,因此引理 2 被排除了。

因此,我们通过将约束 $\lambda_1 = \lambda_2 = \lambda_3 = 1$ 代入式(5.3),得到下面的简化式参数

$$\mathbf{m}(\boldsymbol{\vartheta}) = \left[\frac{\beta_1}{\sqrt{\psi+1}}, \frac{\beta_2}{\sqrt{\psi+1}}, \frac{\beta_3}{\sqrt{\psi+1}}, \frac{\beta_4}{\sqrt{\lambda_4^2\psi+1}}, \frac{\psi}{\psi+1}, \frac{\psi}{\psi+1}, \frac{\lambda_4\psi}{\sqrt{\lambda_4^2\psi+1}\sqrt{\psi+1}}\right.$$

$$\frac{\psi}{\psi+1},\frac{\lambda_4\psi}{\sqrt{\lambda_4^2\psi+1}\sqrt{\psi+1}},\frac{\lambda_4\psi}{\sqrt{\lambda_4^2\psi+1}\sqrt{\psi+1}},\frac{\lambda_4\psi}{\sqrt{\lambda_4^2\psi+1}\sqrt{\psi+1}}\Bigg]$$

将其与式(5.4)等值,但是将截距标记为 $b_i$ 以与参数相区别,用 BPN 参数求解该受约束单因子模型的参数,得

$$\left\{\beta_1=\frac{b_1\sqrt{1-\frac{\rho_1}{-2+\rho_1}}}{\sqrt{2}},\beta_2=\frac{b_2\sqrt{1-\frac{\rho_1}{-2+\rho_1}}}{\sqrt{2}},\beta_3=\frac{b_3\sqrt{1-\frac{\rho_1}{-2+\rho_1}}}{\sqrt{2}},\right.$$

$$\left.\beta_4=\frac{b_4\sqrt{1-\frac{\rho_2^2}{-2\rho_1+\rho_2^2}}}{\sqrt{2}},\lambda_4=\frac{\pm\sqrt{-2+\rho_1\rho_2}}{\sqrt{\rho_1}\sqrt{-2\rho_1+\rho_2^2}},\psi=\frac{-\rho_1}{-2+\rho_1}\right\}$$

该解为唯一解,除了 $\lambda_4$ 的符号以外,该 BPN 模型显然也能在受约束单因子模型的参数空间上生成参数。

用受约束单因子模型的参数来对 BPN 模型参数求解,可得

$$\left\{b_1=\frac{\sqrt{2}\beta_1}{\sqrt{1+\psi}},b_2=\frac{\sqrt{2}\beta_2}{\sqrt{1+\psi}},b_3=\frac{\sqrt{2}\beta_3}{\sqrt{1+\psi}},b_4=\frac{\sqrt{2}\beta_4}{\sqrt{1+\psi\lambda_4^2}}\right.$$

$$\left.\rho_1=\frac{2\psi}{1+\psi},\rho_2=\frac{2\psi\lambda_4}{\sqrt{1+\psi}\sqrt{1+\psi\lambda_4^2}}\right\}$$

因子方差 $\psi$ 显然是非负数,$\psi\geqslant 0$。根据 $\rho_1=2\psi/(1+\psi)$,可得 $\rho_1\geqslant 0$;该受约束单因子模型不能对 BPN 模型生成负的 $\rho_1$。因此,该 BPN 模型与受约束单因子模型不是全局等价。

重要的是,只要约束 $\rho_1\geqslant 0$ 是合理的,我们就可利用等价来大大简化对 BPN 模型的估计。为了获得边缘似然,不再需要去对 4 个随机效应求积分,相反,我们仅需计算一个单维积分(Rabe-Hesketh and Skrondal,2001)。

缺乏全局等价的另一个例子,或许也是更加突出的例子,为带有复合对称残协方差矩阵(compound symmetric residual covariance Matrix)的多元线性模型与随机截距模型。前者将方差设定为等价于 $A$,且协方差等价于 $B$。而对随机截距模型来说,其所暗含的边缘方差和协方差就分别为 $\psi+\theta_{ii}$ 和 $\psi$。注意,由于 $\psi\geqslant 0$,在随机截距模型中协方差必须为非负数;而在复合对称下,协方差 $B$ 并不需要为正数(另见 Lindsey,1999)。

### 5.3.3 经验等价

与对辨识的分析类似,我们现在来考虑关于等价的"经验"分析:

- 对于一个样本,如果两个模型中存在一对一的和参数估计相关的函数,使得似然值几乎相等,那么这两个模型为**经验等价**。

与经验辨识类似,对经验等价的分析基于参数估计,因此不如"理论"等价分析那么形式化。然而,在操作中,对于那些无法用一阶矩和二阶矩来完整表征的模型来说,由于理论方法不可行,我们只能求诸于经验等价。

**BLN 模型及受约束二项单因子模型**

我们现在分析已辨识的 BLN 模型与受约束单因子对数概率模型的经验等价。这两个模型的偏差分别为 6.28 和 6.58。由于偏差非常接近,这两个模型所暗含的一阶矩和

二阶矩几乎相同:所估计的均值分别为 $\hat{\mu}_1 = -0.91, -0.87, \hat{\mu}_2 = -0.99, -0.96, \hat{\mu}_3 = -1.05, -1.03, \hat{\mu}_4 = -1.03, -0.97$,潜响应相关系数估计分别为 $\hat{\rho}_{12} = \hat{\rho}_{13} = \hat{\rho}_{22} = 0.36, 0.34$ 与 $\hat{\rho}_{14} = \hat{\rho}_{24} = \hat{\rho}_{34} = -0.21, -0.21$。

使用下述方程,我们可将单因子模型的参数估计转换为 BLN 模型的估计,即

$$b_r = \beta_r \frac{\sqrt{\sigma^2 + \pi^2/3}}{\sqrt{\psi + \pi^2/3}}, \qquad r = 1,2,3$$

$$b_4 = \beta_4 \frac{\sqrt{\sigma^2 + \pi^2/3}}{\sqrt{\lambda_4^2\psi + \pi^2/3}}$$

$$\rho_1 = \frac{\psi(\sigma^2 + \pi^2/3)}{\sigma^2(\psi + \pi^2/3)}$$

$$\rho_2 = \frac{\lambda_4\psi(\sigma^2 + \pi^2/3)}{\sigma^2\sqrt{(\psi + \pi^2/3)(\lambda_4^2\psi + \pi^2/3)}}$$

其中,我们将 $\sigma$ 替换为最大似然估计值 4.06。所得到的估计(-3.87, -4.27, -4.55, -4.30, 0.41, -0.25)非常接近于 BLN 模型的最大似然估计(-4.04, -4.42, -4.69, -4.56, 0.43, -0.25)。

## 5.4 总结与进一步阅读

我们对辨识和等价概念进行了定义,并且证明这两个属性既可采用解析性分析也可采用经验法分析。通过计算机代数,可以获得雅可比行列式的秩及其零空间的基。为了简单起见,我们通过直接代入雅可比行列式来进行参数约束。或者,我们可以用一个约束矩阵的雅可比行列式来展开本章所用的雅可比行列式(如 Rothenberg,1971)。

我们的讨论限制在参数模型的辨识问题上。更令人感到畏惧的是"非参数"模型的辨识,这些模型由对(未参数化)函数的约束来表征。这种情况下,辨识更加关注对于观测是否存在多个能产生相同分布的函数集。在计量经济学中,讨论了很多包含潜变量的比例风险模型的非参辨识(如 Brinch,2001;Van den Berg,2001)。

到目前为止,我们仅仅讨论了基本参数 $\boldsymbol{\vartheta}$ 为未知常数的模型的辨识问题。在贝叶斯研究领域,切奇利(Chechile,1977)讨论了"后验概率辨识(posterior-probabilistically identified)"这个概念并证明这样的模型可被辨识,尽管它们无法"似然辨识(likelihood-identified)"。关于用先验分布来辨识模型的例子可参见 Knorr-Held and Best (2001)。

值得一提的是,用马尔可夫链蒙特卡洛法或其他模拟方法(见第 6 章)用辨识的观点来看是存在风险的。正如齐内(Keane,1992:193)所指出的,这是由于"模拟误差会产生轮廓线,而真实的目标函数则是扁平的,并且会产生非奇异的海森矩阵,而真实的海森矩阵却是奇异的"。霍洛维茨等(Horowitz,1982)没能发现一个特殊的多项正态概率模型其实是不可辨识的,齐内通过引证霍洛维茨的例子说明了该危险性。

对于本书所讨论的潜变量模型来说,等价并不是一个重要的问题。例如,麦克拉姆等(MacCallum,1993)指出塔克尔(Tucker,1972)所提出的针对三向临近数据(three-way proximity data)的多维尺度模型与卡洛尔和张(Carroll and Chang,1972)的模型其实是等价的,尽管它们在判断任务上对个体差异的表现非常不同。等价也可设计出不同类型的模型,如潜类模型和拉什模型(如 Lindsay et al.,1991;Heinen,1996)或连续响应的因子模型与潜剖面模型(如 Bartholomew,1987,1993;Molenaar and von Eye,1994)。

贝克尔等(Bekker,1994)对于参数模型的辨识和等价问题做了全新的且非常全面的讨论,包括形式定义、假设及定理。关于辨识的一些有所助益的处理可参见 Koopmans and Reiersøl (1950)、Wald (1950)、Anderson and Rubin (1956)、Fisher (1966)、Geraci (1976)、Rothenberg(1971)、Dupacôvá and Wold(1982)、Hsiao(1983)、Rabe-Hesketh and Skrondal (2001)及 Bechger et al. (2001)。对于等价研究的贡献还包括 Stelzl(1986)、Breckler(1990)、Jöreskog and Sörbom(1990)、Luijben(1991)、McCallum et al. (1993)、Hershberger(1994)、Raykov and Penev (1999)、Rabe-Hesketh and Skrondal (2001)及 Bechger et al. (2002)。

在计量经济学中,由于复杂的结构模型已使用了很长时间,因此,对于辨识问题给予了大量关注。与此相反,在生物计量学中,传统上使用更加简单的模型。在计量心理学中,等价问题看起来得到越来越多的关注。然而,随着高度结构化的模型的使用越来越普遍,辨识与等价问题应该毫无疑问在统计上被给予更多的关注。

# 6 估 计

## 6.1 引 言

本章介绍一些第 4 章提出的属于一般模型框架下的潜变量模型的估计方法。既然它们中的某些方法属于特殊的方法论范畴，因此，我们认为，对这些不同方法做些相关的非技术性的综述是必要的。本章附录中对这些不同估计方法的软件实现做了一个不甚完整的概述。

我们还对不同方法的优缺点进行了考虑。模型的评判标准有基于通用性、稳健性、计算效率、对缺失数据的处理、估计量的计算等，然而这些估计方法的表现却是参差不齐的。

到目前为止在这本书中，我们将潜变量当做是随机的，而参数则是固定的，这是最为常见的做法。除此之外，无论是出于理论上的或计算方便的原因，潜变量和参数两者都可以被视为固定的。与此相反，贝叶斯学派则将潜变量和参数都处理为随机变量。虽然贝叶斯通常被看做是根本不同的统计范式，但由于实际的原因，这一做法目前还往往被采用。认识到这些不同的观点对于介绍各种不同的估计方法是非常重要的。

**随机潜变量和固定参数**

当潜变量是随机的而参数是固定时，推论通常基于**边缘似然**(marginal likelihood)，即源自潜变量数据的似然，对潜变量分布求积分(对于离散情形，则是求和)。对于连续潜变量的情况，似然通常没有闭式。因此，在 6.3 节中，我们将介绍一些或多或少准确的积分近似方法，包括数值积分和蒙特卡洛积分(模拟似然)。在 6.4 节中，评述了求最大似然的不同方法，包括 EM 算法和牛顿-拉夫逊算法。

在 6.5 节，我们讨论**非参最大似然估计**(nonparametric maximum likelihood estimation, NPMLE)，我们在那里释放了正态潜变量假设。**受约束最大似然**(restricted maximum likelihood, REML)的观点在 6.6 节做简要介绍。对于某些模型，使用 6.7 节中所描述的**有限信息**(limited information)方法，积分维度可以大大地减少。6.8.3 节介绍**罚伪似然方法**(penalized quasi-likelihood, PQL)，即一个避免积分的近似方法，在 6.9 节讨论类似的**广义估计方程**(generalized estimating equations, GEE)算法。GEE 非常不同于本书中涉及的其他方法，因为响应之间的相关性并不用潜变量显性建模，而是将其视为扰动。此外，回归参数也不再可解释为条件效应或特定群组的效应，而是解释为边缘效应或总体平均效应。

**固定潜变量和参数**

当潜变量被视为未知的**固定参数**，而不是随机变量，则积分可以避免。固定效应方法可被视为条件依赖于样本中的效应。在这种情况下，群组是否实际被视为总体的一个随机样本，已经无关紧要了。在6.10节中，我们介绍两种固定效应估计方法。在**联合最大似然**(joint maximum likelihood，JML)估计中，潜变量和模型参数是联合估计的，而在**条件最大似然**(conditional maximum likelihood，CML)估计中，潜变量宽泛地被称为"条件外"。

**随机潜变量和参数**

在6.11节中所介绍的贝叶斯方法，将潜变量和参数都当做随机的，且推论基于观测数据的**后验分布**。在6.11.5节中，我们描述比较流行的马尔可夫链蒙特卡洛(MCMC)方法，此方法用于从后验分布中抽样以及用它们的后验均值估计参数。

## 6.2 最大似然：封闭形式的边缘似然性

边缘似然性中的积分在某些情况下可以被显性求解并用闭式表达，典型的例子有LISREL模型和线性混合模型。在这些情况下，来源于潜变量的多元正态潜变量和多元正态响应将产生**多元正态边缘分布**(multivariate normal marginal distributions)。

线性混合模型的估计将在6.8.1节讨论。对于3.5节所介绍的LISREL模型，模型隐含的协方差矩阵可被表示为

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}(\mathbf{I}-\mathbf{B})^{-1}\boldsymbol{\Psi}(\mathbf{I}-\mathbf{B})^{-1}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}$$

在这些模型中，通常不关心均值结构，因而，在这一节中，我们令 $n$ 维响应向量 $\mathbf{y}_j$ 的期望值为零。则似然可表示为

$$f(\boldsymbol{\vartheta};\mathbf{Y}) = \prod_{j=1}^{J}(2\pi)^{-\frac{n}{2}}|\boldsymbol{\Sigma}^{-1}|\exp(-\mathbf{y}_j'\boldsymbol{\Sigma}^{-1}\mathbf{y}_j)$$

$\mathbf{y}$ 的经验协方差矩阵 $\mathbf{S}$ 是对参数结构化 $\boldsymbol{\Sigma}$ 的充分统计。$\mathbf{S}$ 具有威沙特分布(Wishart distribution)，由此可得(如Jöreskog，1967)，我们等价地最小化未知的自由参数 $\boldsymbol{\Lambda}$，$\boldsymbol{\Psi}$ 和 $\boldsymbol{\Theta}$ 的拟合函数

$$F_{ML} = \log|\boldsymbol{\Sigma}| + \mathrm{tr}(\mathbf{S}\boldsymbol{\Sigma}^{-1}) - \log|\mathbf{S}| - n$$

来取代最大化似然。$F_{ML}$非负，并且当完全拟合时，即当所拟合的 $\boldsymbol{\Sigma}$ 等于 $\mathbf{S}$ 时，$F_{ML}$为零。拟合函数还为最大似然估计提供了一个估计信息矩阵。

布朗尼(Browne，1984)为方差结构提出了一组加权最小二乘(WLS)拟合函数，即

$$F_{WLS} = [\boldsymbol{\sigma} - \mathbf{s}]'\mathbf{W}^{-1}[\boldsymbol{\sigma} - \mathbf{s}] \qquad (6.1)$$

其中，$\boldsymbol{\sigma}$ 和 $\mathbf{s}$ 分别是模型隐含的和经验协方差矩阵中的包含非冗余元素的向量，$\mathbf{W}$ 是正定加权矩阵。最大似然估计是通过把 $\hat{\boldsymbol{\Sigma}}$(参数估计所隐含的协方差矩阵)作为权重矩阵来获得的。对于没有闭式的模型的有限信息估计，WLS方法也是有用的，这个我们将在6.7节讨论。

广义线性随机截距模型也可能有闭式似然。具体而言，将产生均值为 $\exp(v)$ 的泊松分布响应与产生均值的伽玛分布相结合，即形成**负二项**(negative binomial)边缘模型(如，Greenwood and Yule，1920；Hausman et al.，1984)。对于二项响应，众所周知，假定概率满足贝塔分布的**贝塔二项**(beta binomial)模型具有闭式似然(如Skellam，1948；Williams，1975)。这里，对边缘对数概率通常设定一个回归模型(如Heckman and Willis，1977)。

与负二项模型不同,这不是一个广义线性混合模型,因为在线性预测中它不能包含附加的随机截距。遗憾的是,这些有用的结果并不适用于常见的不同变量在群组内变化的情况(如 Neuhaus and Jewell,1990)。

## 6.3 最大似然:近似边缘似然

给定解释变量,响应的简化式分布由 4.9.1 节中所述的潜变量积分导出。作为给定响应的基本参数的函数,就是边缘似然函数 $f(\boldsymbol{\vartheta};\mathbf{y},\mathbf{X})$。

对于广义模型,边缘似然为

$$f(\boldsymbol{\vartheta};\mathbf{y},\mathbf{X}) = \prod g^{(L)}(\mathbf{y}_{(L)})$$

其中,乘积遍布所有顶层群组。根据在 $l$ 层及其上层的潜变量,$\boldsymbol{\zeta}^{(l+)}=(\boldsymbol{\zeta}^{(l)\prime},\cdots,\boldsymbol{\zeta}^{(L)\prime})'$,令 $g^{(l)}(\mathbf{y}_{(l)}|\boldsymbol{\zeta}^{(l+)})$ 为 $l$ 层单元的响应的联合条件概率(密度)。从 $l=2$ 开始,我们可以递归估计积分

$$g^{(l)}(\mathbf{y}_{(l)}\mid\boldsymbol{\zeta}^{([l+1]+)}) = \int h^{(l)}(\boldsymbol{\zeta}^{(l)})\prod g^{(l-1)}(\mathbf{y}_{(l-1)}\mid\boldsymbol{\zeta}^{(l+)})\mathrm{d}\boldsymbol{\zeta}^{(l)} \tag{6.2}$$

直到第 $L$ 层。我们设

$$g^{(l)}(\mathbf{y}_{(l)}\mid\boldsymbol{\zeta}^{([l+1]+)}) \equiv g^{(l)}(\mathbf{y}_{(l)}\mid\mathbf{X}_{(L)},\boldsymbol{\zeta}^{([l+1]+)};\boldsymbol{\vartheta})$$

以简化标记,并且在本章其余部分,我们将继续如此简化。

我们将详细描述一些双层次随机截距模型的积分方法,并说明他们是如何推广到一般模型的。令 $\eta_j^{(2)}=\zeta_j$,随机截距模型由下式给出

$$v_{ij} = x'_{ij}\beta + \zeta_j$$

对于第二层次的第 $j$ 个单元,其响应的联合密度为

$$g^{(2)}(\mathbf{y}_{j(2)}) = \int_{-\infty}^{\infty} h(\zeta_j)\prod_i g^{(1)}(y_{ij}\mid\zeta_j)\mathrm{d}\zeta_j \tag{6.3}$$

遗憾的是,这些积分通常没有包含闭式。但有以下几个方法来近似积分:

- 拉普拉斯近似。
- 使用积分法或自适应积分法的数值积分。
- 蒙特卡洛积分。

这些将在 6.3.1 节至 6.3.3 节中阐述。6.3.4 节将描述一种多元正态潜响应的专用模拟方法,此方法基于潜响应积分而不是潜变量积分(见 4.9 节)。

### 6.3.1 拉普拉斯近似

对于单维积分,拉普拉斯近似可写为

$$\begin{aligned}\int_{-\infty}^{\infty}\exp[f(x)]\mathrm{d}x &\approx \int_{-\infty}^{\infty}\exp[f(\tilde{x})-(x-\tilde{x})^2/2\sigma^2]\mathrm{d}x\\ &= \int_{-\infty}^{\infty}\exp[f(\tilde{x})]\sqrt{2\pi}\sigma\phi(x;\tilde{x},\sigma^2)\mathrm{d}x\\ &= \exp[f(\tilde{x})]\sqrt{2\pi}\sigma \end{aligned} \tag{6.4}$$

其中,$\phi(x;\tilde{x},\sigma^2)$ 是一个均值为 $\tilde{x}$,方差为 $\sigma^2$ 的正态密度,$\tilde{x}$ 是 $f(x)$ 的众数,从而也是 exp $[f(x)]$ 的众数,且

$$\sigma^2 = -\left(\frac{\partial^2 f(\tilde{x})}{\partial x^2}\right)^{-1}$$

即函数$f(x)$对$\tilde{x}$的二阶导数的逆负函数在众数$\tilde{x}$处的值。

近似由$\tilde{x}$在其众数处的二阶泰勒级数展开而得,因此,一阶项消失(见式(6.4)第一行括号内)。如果被积函数与拥有均值$\tilde{x}$和方差$\sigma^2$的正态密度成比例,则此近似是精确的,因为在此情况下,$f(x)$是$x$的积分。

对于一个随机截距模型,我们需要估计式(6.3)中的积分。被积函数(关于 exp$[f(x)]$)

$$h(\zeta_j)\prod_i g^{(1)}(y_{ij}\mid\zeta_j) \tag{6.5}$$

是$\zeta_j$的先验密度和给定$\zeta_j$条件下响应的联合概率(密度)的产物。对$\zeta_j$归一化后,被积函数则成为给定群组$j$的观测响应条件下,$\zeta_j$的后验密度(另见7.2节)。因而,在拉普拉斯近似中,$\tilde{x}$对应于后验众数$\tilde{\zeta}_j$,$\sigma$则对应于处于众数$\sigma_j$的后验的曲率。近似即变为

$$\begin{aligned}\ln g^{(2)}(\mathbf{y}_{j(2)}) &\approx \ln(\sqrt{2\pi}\sigma_j)+\ln h(\tilde{\zeta}_j)+\sum_i \ln g^{(1)}(y_{ij}\mid\tilde{\zeta}_j)\\ &=\ln(\sigma_j/\sqrt{\psi})-\tilde{\zeta}_j^{\,2}/(2\psi)+\sum_i \ln g^{(1)}(y_{ij}\mid\tilde{\zeta}_j)\end{aligned} \tag{6.6}$$

当随机截距的后验密度近似正态时,该近似效果很好。众所周知,这是在大样本(此处指群组规模)的情况下。此渐近归一化有时涉及贝叶斯中心极限定理(如,Carlin and Louis,1998)。当条件响应概率更加趋于正态时,后验也更加趋于正态,如具有大均值的泊松或具有大分母的二项分布。在此情况下,后验众数接近后验均值,同时$\sigma_j$接近后验标准差。

对于罚伪似然(PQL)方法(请参见 Shall,1991;McGilchrist,1994;Breslow and Clayton,1993),式(6.6)中的第一项被忽略,并且使所剩余的项关于固定效应参数$\boldsymbol{\beta}$(已知的变量参数)最大化。这里有必要注意的是,这与最大似然并不相关。取而代之的是,罚伪对数似然

$$-\zeta_j^2/(2\psi)+\sum_i \ln g^{(1)}(y_{ij}\mid\zeta_j)$$

是连带最大化的。毫无疑问,既然$\tilde{\zeta}_j$使式(6.5)最大且式(6.5)的对数不同于上述的地方仅仅在于常数$\ln(\sqrt{2\pi\psi})$,则其关于$\boldsymbol{\beta}$以及关于给定$\boldsymbol{\beta}$的$\zeta_j$必然最大化。PQL 方法的另一种推导将在6.8.3节中讨论。

李和奈尔德(Lee and Nelder,1996,2001)定义了**分级似然**(hierarchical likelihood),或称$h$-似然,即响应变量与潜变量的联合分布,其中将潜变量当做观测变量。因此,$h$-似然的对数为

$$l_h=-\zeta_j^2/(2\psi)+\sum_i \ln g^{(1)}(y_{ij}\mid\zeta_j)-\ln(\sqrt{2\pi\psi}) \tag{6.7}$$

使$h$-似然关于$\boldsymbol{\beta}$和$\zeta_j$最大化(对于固定的$\psi$)将得到和罚伪似然一样的估计。$h$-似然的优点在于它并不要求积分并且允许灵活设定潜变量分布。尽管如此,在缺失数据问题中,雷托和鲁宾(Little and Rubin,1983;2002:124)认为这种方法不能"广泛地共享 ML 估计的优化属性,除非在缺失数据的比例随样本规模增加而趋近于0的情况下平凡渐近"。在潜变量模型中,缺失数据是潜变量的现实化,因此,只有当群组规模趋近于无穷时,缺失数据的比例才趋近于0;另见克雷顿(Clayton,1996a)。李和奈尔德(Lee and Nelder,1996)也对此进行了有益的讨论。

对于线性混和模型,$h$-似然的似然方程即是著名的对于$\boldsymbol{\beta}$和$\boldsymbol{\zeta}$的"混和模型方程"

(Henderson,1975;Harville,1977)。对于给定的随机效应协方差矩阵 $\boldsymbol{\Psi}$,对 $\boldsymbol{\beta}$ 的估计量是最大边缘似然估计量,并且对 $\boldsymbol{\zeta}$ 的估计量是经验贝叶斯预测量或者 7.3.1 节中所讨论的最佳线性无偏预测量(BLUP)。

斯特雷特里等人(Stiratelli et al.,1984)针对包含二项响应的随机效应对数概率回归模型,通过在 $\boldsymbol{\beta}$ 的扩散先验条件下,使 $\boldsymbol{\beta}$ 和 $\zeta_j$ 的后验分布最大化(这样后验本质上仍是 $h$-似然),推导出了同样的方程。

拉普拉斯近似基于对 $f(x)$ 的二阶泰勒级数展开。若后验密度非正态并且用二阶拉普拉斯所获得的参数估计来校正偏差,则四阶拉普拉斯近似更为精确(Breslow and Lin,1995;Lin and Breslow,1996)。劳登布什等人(Raudenbush et al.,2000)提出了采用六阶拉普拉斯近似的近似最大似然,即所谓的 LaPlace6。

在对二项响应的小型模拟研究中,六阶拉普拉斯近似显著的优于 PQL,些微的好于 20-点高斯-埃尔米特积分(Gauss-Hermite quadrature)(Ruadenbush and Yang,1998;Raudenbush et al.,2000),与 7-点自适应积分类似。尽管如此,Laplace6(如在 HLM 中被采用)计算会显著的快于自适应积分(如在 SAS NLMIXED 中被采用)。据我们所知,这种方法迄今为止只能被用于具有嵌套随机效应的广义线性混和模型中。

## 6.3.2 数值积分

**高斯-埃尔米特积分**

该积分法对在一组值或被积出来的变量处的被积函数加权求和来做为积分近似。其位置和权值称为**积分法则**(quadrature rules)。当 $f(x)$ 是一个 $(2R-1)$ 次多项式时,高斯-埃尔米特积分法则用以计算如下形式的具有 $R$ 个点的积分

$$\int_{-\infty}^{\infty} \exp(-x^2) f(x)\,\mathrm{d}x \approx \sum_{r=1}^{R} p_r^* f(a_r^*)$$

由于"加权函数"$\exp(-x^2)$ 与正态密度成比例,因此,我们可用该法则来积分出式(6.3)中的正态分布潜变量。首先,我们将积分变量变为标准正态变量 $v_j = \zeta_j/\sqrt{\psi}$,使得式(6.3)中的积分变为

$$g^{(2)}(\mathbf{y}_{j(2)}) = \int_{-\infty}^{\infty} \phi(v_j) \prod_i g^{(1)}(y_{ij} \mid \sqrt{\psi} v_j)\,\mathrm{d}v_j \tag{6.8}$$

其中,$\phi(\cdot)$ 为标准正态密度

$$\phi(v_j) = \frac{1}{\sqrt{2\pi}} \exp(-v_j^2/2)$$

将高斯-埃尔米特积分法则应用于该积分,则可得到

$$\int_{-\infty}^{\infty} \phi(v_j) \prod_i g^{(1)}(y_{ij} \mid \sqrt{\psi} v_j)\,\mathrm{d}v_j \approx \sum p_r \prod_i g^{(1)}(y_{ij} \mid \sqrt{\psi} a_r) \tag{6.9}$$

其中,$p_r \equiv p_r^*/\sqrt{\pi}$,以及 $a_r \equiv \sqrt{2} a_r^*$。

式(6.2)中的多变量积分可以用**笛卡儿乘积积分**(cartesian product quadrature)求得。这里我们将积分变量变为独立正态分布潜变量 $\mathbf{v}^{(l)}$,从而

$$\boldsymbol{\zeta}^{(l)} = \mathbf{Q}^{(l)} \mathbf{v}^{(l)} \tag{6.10}$$

其中,$\mathbf{Q}^{(l)}$ 为 $\boldsymbol{\zeta}^{(l)}$ 的协方差矩阵 $\boldsymbol{\Psi}^{(l)}$ 的乔莱斯基分解。在第 $l$ 层对 $M_l$ 个潜变量积分则变为

$$g^{(l)}(\mathbf{y}_{(l)} \mid \mathbf{v}^{([l+1]^+)})$$

$$= \int_{-\infty}^{\infty} \phi(v_1^{(l)}) \cdots \int_{-\infty}^{\infty} \phi(v_{M_l}^{(l)}) \prod g^{(l-1)}(\mathbf{y}_{(l-1)} \mid \mathbf{v}^{(l+)}) \mathrm{d}v_{M_l}^{(l)} \cdots \mathrm{d}v_1^{(l)}$$

$$\approx \sum_{r_1=1}^{R_1^{(l)}} p_{r1} \cdots \sum_{r_{M_l}=1}^{R_{M_l}^{(l)}} p_{rM_l} \prod g^{(l-1)}(\mathbf{y}_{(l-1)} \mid a_{r1}, \cdots, a_{rM_l}, \mathbf{v}^{([l+1]^+)}) \quad (6.11)$$

其中，$\mathbf{v}^{(l+)} = (\mathbf{v}^{(l)\prime}, \cdots, \mathbf{v}^{(L)\prime})'$。

因此，潜变量可由如图 6.1 所示的左上角矩形网格点计算而得。不同的积分点 $R_m^{(l)}$ 可用于不同的潜变量 $v_m^{(l)}$，整个积分共需要 $\prod_{m=1}^{M_l} R_m^{(l)}$ 次计算。

另一种取代笛卡儿乘积积分法则的方法是**球面积分**(spherical quadrature)，专用于对多元正态密度(Stroud，1971)。正如其名字所述，积分点呈(超)球面状(二维圆)排列而不是矩形，如图 6.1 所示的左下图。重要的是，在给定所获精度下，球面法则比笛卡儿法则所需点少。尽管如此，在一些维数和要求的精度下，目前没有可用的球面法则。这或许是因为除了 Clarkson and Zhan(2002)以及 Rabe-Hesketh et al. (2004b)之外，求解法则还未被广泛的用于潜变量模型。

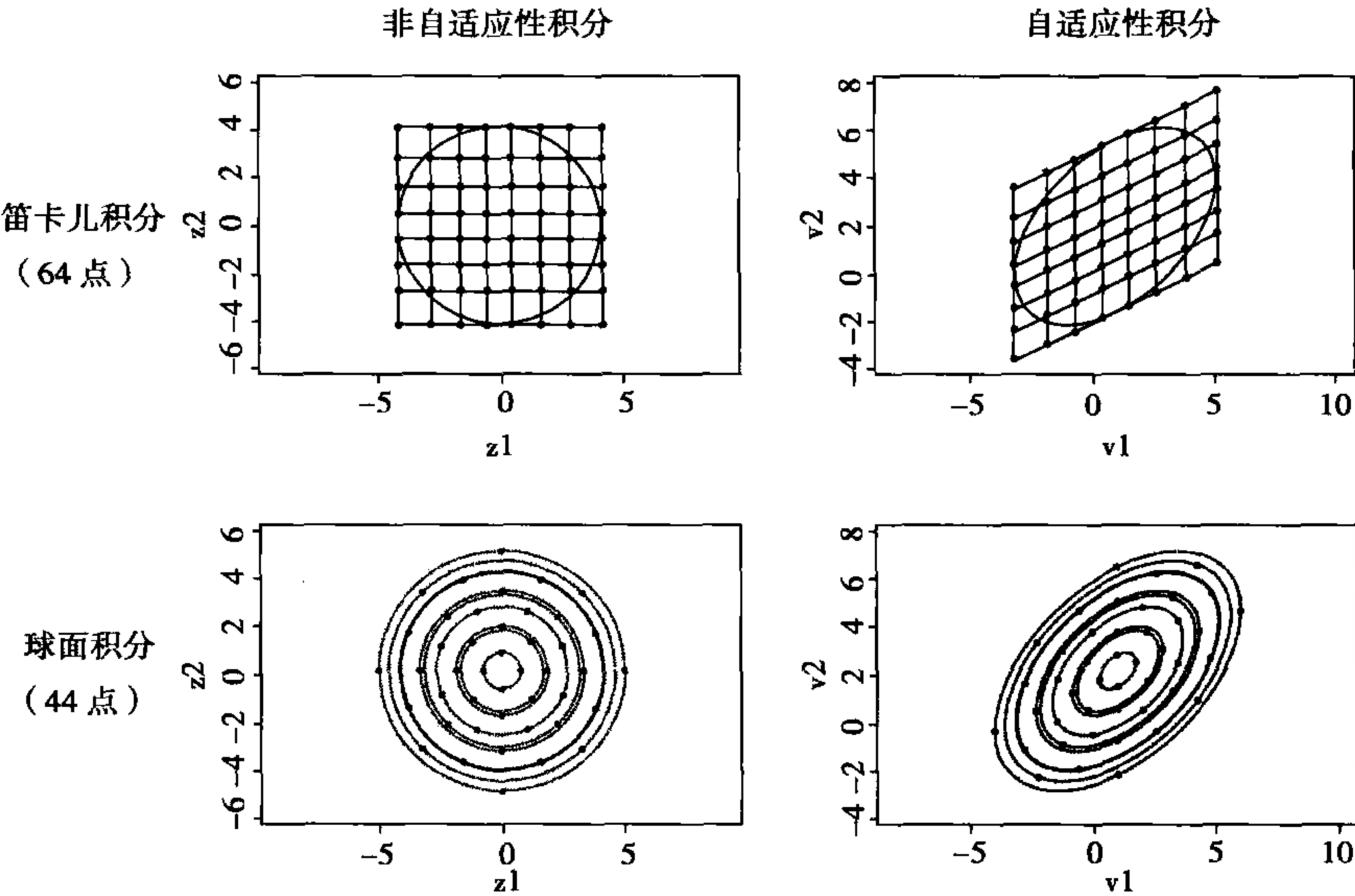



**图 6.1 使用笛卡儿乘积积分与球面积分法则的非自适应积分和自适应积分的原点，其中，$\mu_1 = 1, \mu_2 = 2, \tau_1 = \tau_2 = 1$，后验相关为 0.5**

(资料来源：Rabe-Hesketh et al.,2004b)

对于正态概率题器响应模型(IRT)，博科和李博曼(Bock and Lieberman，1970)使用费舍尔计分算法的高斯积分(Gaussian quadrature)(见 6.4.2 节)；而博科和艾特金(Bock and Aitkin，1981)则使用 EM 算法的高斯积分(见 6.4.1 节)。巴特勒和墨菲特(Butler and Moffitt，1982)提出针对随机截距正态概率回归模型的积分。高斯积分也可用于其他模型，如单参数对数概率 IRT 模型(Thissen，1982)、广义线性混和模型(如 Hedeker，Gibbons，1994，1996a)。

当式(6.9)的乘积可用低阶多项式拟合的较好时，高斯积分可达到很好的效果。尽管如此，实际上经常需要大量的积分点来近似然(如 Crouch and Speiengelman，1990)，如在一个群组 $j$ 中存在大量的第一层次单元这种情况。式(6.9)中的乘积将会存在一个非常尖的峰值，很难用一个多项式来近似。如果使用的积分点不足，峰值会位于自适应积分点 $\alpha_r$ 和 $\alpha_{r+1}$ 之间，因而群组 $j$ 的大部分似然贡献将丢失(见图 6.2 中的上半部分)。勒

斯弗和斯皮森斯(Lesaffre and Spiessens,2001)在求解二项响应、阿尔伯特和弗曼(Albert and Follmann,2000)在处理计数响应时,都提出了这些问题。注意,对于计数响应,如果个体 $g^{(1)}(y_{ij}|v_j)$ 具有一个尖的峰值,那么积分近似即便在小规模的群组中也会失效。如果存在大的随机效应方差,近似也同样失效。小模拟研究表明,高斯-埃尔米特积分优于PQL(Raudenbush and Yang,1998;Raudenbush et al.,2000),但劣于自适应积分(Rabe-Hesketh et al.,2004b)。

**自适应积分**

为克服普通积分的问题,自适应积分本质上将积分位置平移和比例缩放到被积函数峰值以下。正如6.3.1节中讨论的那样,被积函数

$$\phi(v_j)\prod_i g^{(1)}(y_{ij}\mid v_j)$$

与先验密度成比例,且通常能够用具有群组特定均值 $\mu_j$ 和方差 $\tau_j^2$ 的正态密度 $\phi(v_j;\mu_j,\tau_j^2)$ 来较好地近似。与式(6.9)中应用积分法则时,将**先验**密度当做"加权函数"不同,我们可以重写积分为

$$g^{(2)}(\mathbf{y}_{j(2)}) = \int_{-\infty}^{\infty}\phi(v_j;\mu_j,\tau_j^2)\left[\frac{\phi(v_j)\prod_j g^{(1)}(y_{ij}\mid\sqrt{\psi}v_j)}{\phi(v_j;\mu_j,\tau_j^2)}\right]\mathrm{d}v_j \tag{6.12}$$

并将近似**后验**密度的正态密度作为加权函数。

将积分变量从 $v_j$ 变为 $z_j=(v_j-\mu_j)/\tau_j$ 并应用标准积分法则,可得

$$\begin{aligned}
g^{(1)}(\mathbf{y}_{j(2)}) &= \int_{-\infty}^{\infty}\frac{\phi(z_j)}{\tau_j}\left[\frac{\phi(\tau_j z_j+\mu_j)\prod_i g^{(1)}(y_{ij}\mid\sqrt{\psi}(\tau_j z_j+\mu_j))}{\frac{1}{\sqrt{2\pi}\tau_j}\exp(-z_j^2/2)}\right]\tau_j\mathrm{d}z_j\\
&\approx \sum_{r=1}^{R}p_r\left[\frac{\phi(\tau_j a_r+\mu_j)\prod_i g^{(1)}(y_{ij}\mid\sqrt{\psi}(\tau_j a_r+\mu_j))}{\frac{1}{\sqrt{2\pi}\tau_j}\exp(-a_r^2/2)}\right]\\
&= \sum_{r=1}^{R}\pi_{jr}\prod_i g^{(1)}(y_{ij}\mid\sqrt{\psi}\alpha_{jr})
\end{aligned}$$

其中

$$\alpha_{jr}\equiv\tau_j a_r+\mu_j$$

且

$$\pi_{jr}\equiv\sqrt{2\pi}\tau_j\exp(a_r^2/2)\phi(\tau_j a_r+\mu_j)p_r$$

当后验密度近似正态时,方括号里面的项可用低阶多项式较好的近似,因此,分子与分母近似成比例。可以想象,这种方法所需要的积分点比非自适应积分更少,并且对大规模群组有效。自适应积分的有效性如图6.2所示,说明当 $R=5$ 时,自适应积分如何平移以及比例缩放积分位置故而直接排列在被积函数之下。

当存在多个潜变量时,后验协方差必须考虑。详见Naylor and Smith(1988)以及Rabe-Hesketh et al.(2004b)。对于具有后验相关系数0.5的两个潜变量,图6.1的第二列表明了自适应积分如何变换以更接近于(近似)二元正态后验的椭圆周线。

奈勒和斯密斯(Naylor and Smith,1982)用近似后验密度的正态密度的均值 $\mu_j$ 和方差 $\tau_j^2$ 作为后验均值和方差。遗憾的是,这些后验矩完全未知并且它们又需要用自适应积分来计算。因此积分反复迭代。用初始值 $\mu_j^0=0$ 和 $\tau_j^0=1$ 定义 $\alpha_{jr}^0$ 和 $\pi_{jr}^0$,第 $k$ 次迭代的后验均值和方差可用下式得到

$$g^{(2)k}(\mathbf{y}_{j(2)}) = \sum_{r=1}^{R} \pi_{jr}^{k-1} \prod_{i=1}^{n_j} g^{(1)}(y_{ij} \mid \sqrt{\psi}\alpha_{jr}^{k-1})$$

$$\mu_j^k \approx \sum_{r=1}^{R} \alpha_{jr}^{k-1} \left[ \frac{\pi_{jr}^{k-1} \prod_{i=1}^{n_j} g^{(1)}(y_{ij} \mid \sqrt{\psi}\alpha_{jr}^{k-1})}{g^{(2)k}(\mathbf{y}_{j(2)})} \right]$$

$$(\tau_j^k)^2 = \sum_{r=1}^{R} (\alpha_{jr}^{k-1})^2 \left[ \frac{\pi_{jr}^{k-1} \prod_{i=1}^{n_j} g^{(1)}(y_{ij} \mid \sqrt{\psi}\alpha_{jr}^{k-1}}{g^{(2)k}(\mathbf{y}_{j(2)})} \right] - (\mu_j^k)^2$$

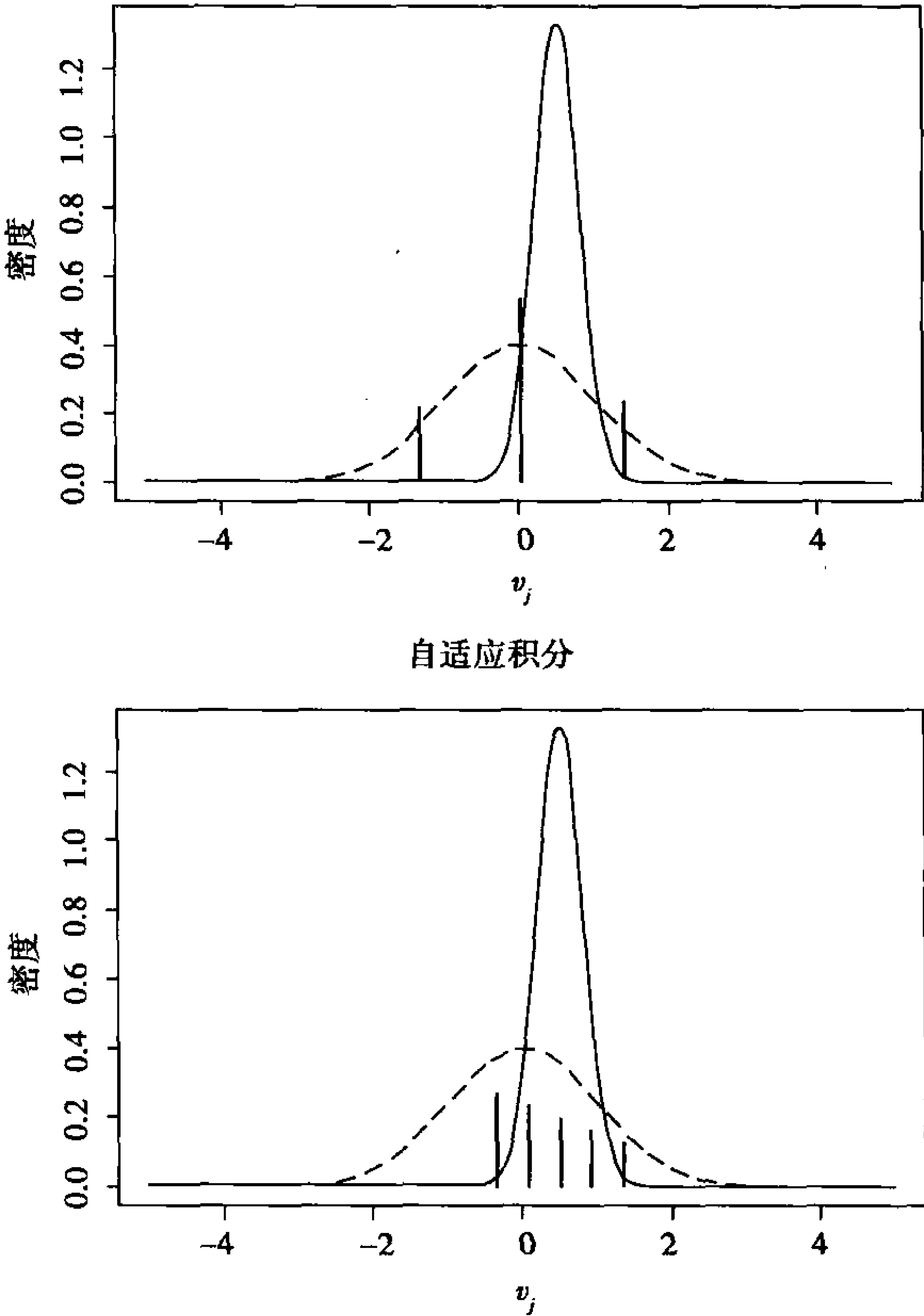



**图 6.2 普通积分和自适应积分的先验密度(点线)、后验密度(实线)及积分权重(条形图)。注意,被积函数与后验密度成比例。**

(资料来源:Rabe-Hesketh et al.,2002)

重复上述步骤直至收敛。奈勒和斯密斯(Naylor and Smith,1988)介绍了一种类似的迭代方法。

另一种计算后验矩 $\mu_j$ 和方差 $\tau_j^2$ 的方法为用众数和该众数下的曲率(Liu and Pierce,1994)作为6.3.1节中所介绍的一阶拉普拉斯近似。在此情况下,采用 $R=1$ 个积分点的自适应积分等价于一阶拉普拉斯近似。使用众数和该众数下的曲率来取代后验矩的优点在于,前者的计算不要求数值积分。尽管如此,该方法不容易推广为多层次模型,正因为这一点,拉贝-赫斯基思等(Rabe-Hesketh et al.,2004b)采纳奈勒和斯密斯的方法来计算广义模型框架下的模型。

皮涅罗和贝茨(Pinheiro and Bates,1995)将自适应积分法用于双层次非线性混和模型,博科和希林(Bock and Schilling,1997)将其应用于包含二项响应的探索性因子分析,拉贝-赫

斯基思等(Rabe-Hesketh et al.,2002,2004b)则将它用于广义线性潜混和模型。自适应积分,即 Stata 程序 gllamm 中所实现的方法,应用于本书大多数包含连续潜变量的例子中。

针对包含二项响应,并且群组规模不同、组内相关也不同的双层次模型,有人做了一个蒙特卡洛实验,以比较自适应积分和普通积分的性能(Rabe-Hesketh et al.,2004b)。自适应积分的性能非常好,所需的积分点少于普通积分。对于大规模群组以及高度组内相关的情形,普通积分有时会失效,而自适应积分只要有足够的积分点就能运作良好。

### 6.3.3 蒙特卡洛积分

令 $\boldsymbol{\varphi}$ 为具有分布 $h(\boldsymbol{\varphi})$ 的随机变量向量。假定我们需要关于 $\boldsymbol{\varphi}$ 的函数 $f(\boldsymbol{\varphi})$ 的期望,则

$$\mathrm{E}[f(\boldsymbol{\varphi})] = \int_{-\infty}^{\infty} h(\boldsymbol{\varphi})f(\boldsymbol{\varphi})\mathrm{d}\boldsymbol{\varphi}$$

蒙特卡洛积分近似于关于 $\boldsymbol{\varphi}$ 的模拟值的函数 $f(\boldsymbol{\varphi})$ 的期望均值。不同的模型方法,产生不同形式的蒙特卡洛积分。

**原始蒙特卡洛积分**

在此情况下,给定模拟量

$$\mathrm{E}[f(\boldsymbol{\varphi})] \approx \bar{f} \equiv \frac{1}{R}\sum_{r=1}^{R} f(\boldsymbol{\varphi}^{(r)})$$

从 $h(\boldsymbol{\varphi})$ 抽取独立样本 $\boldsymbol{\varphi}^{(r)}, r=1,\cdots,R$。

由大数定理可得,当 $R\to\infty$ 时,$\bar{f}$ 依概率 1 收敛于 $\mathrm{E}(f(\boldsymbol{\varphi}))$。根据中心极限定律,当 $R$ 大到均值为 $\mathrm{E}(f(\boldsymbol{\varphi}))$ 及方差为

$$\mathrm{Var}(\bar{f}) = \frac{1}{R(R-1)}\sum_{r=1}^{R}[f(\boldsymbol{\varphi}^{(r)}) - \bar{f}]^2$$

则 $\bar{f}$ 近似正态分布。

令 $\boldsymbol{\varphi} = v_j, h(\boldsymbol{\varphi}) = \phi(\boldsymbol{\varphi})$ 且 $f(\boldsymbol{\varphi}) = \prod_i g^{(1)}(y_{ij} \mid \sqrt{\psi}v_j)$,随机截距模型的似然

$$\int_{-\infty}^{\infty} \phi(v_j)\prod_i g^{(1)}(y_{ij} \mid \sqrt{\psi}v_j)\mathrm{d}v_j$$

其形式为 $\mathrm{E}(f(\boldsymbol{\varphi}))$。似然的蒙特卡洛积分可通过从 $\phi(v_j)$ 抽样,并计算均值 $\prod_i g^{(1)}(y_{ij} \mid \sqrt{\psi}v_j)$ 来进行。

仅当群组规模 $n_j$ 增加时,拉普拉斯近似性能才能获得改进;与此不同的是,在蒙特卡洛积分中,我们可以简单地通过增加模拟次数 $R$ 来提高精度。此外,与积分法或拉普拉斯近似不同,我们可根据估计模拟量的方差来评估近似精度。

**重点抽样**

原始蒙特卡洛积分可用重点抽样来减小抽样方差,从而获得改进。当难以从 $h(\boldsymbol{\varphi})$ 中抽取 $\boldsymbol{\varphi}$ 或 $h(\boldsymbol{\varphi})$ 不平滑时,估计 $\mathrm{E}(f(\boldsymbol{\varphi}))$ 需要审慎选择重点密度 $g(\boldsymbol{\varphi})$,则积分可写为

$$\mathrm{E}[f(\boldsymbol{\varphi})] = \int_{-\infty}^{\infty} g(\boldsymbol{\varphi})\frac{h(\boldsymbol{\varphi})f(\boldsymbol{\varphi})}{g(\boldsymbol{\varphi})}\mathrm{d}\boldsymbol{\varphi}$$

其中,$g(\boldsymbol{\varphi})$ 是这样的密度:(a)容易抽取 $\boldsymbol{\varphi}$;(b)支集与 $f(\boldsymbol{\varphi})$ 相同;(c)给定 $\boldsymbol{\varphi}$ 容易估计 $\frac{h(\boldsymbol{\varphi})f(\boldsymbol{\varphi})}{g(\boldsymbol{\varphi})}$;(d)在 $\boldsymbol{\varphi}$ 的支集的参数范围内,$\frac{h(\boldsymbol{\varphi})f(\boldsymbol{\varphi})}{g(\boldsymbol{\varphi})}$ 有界且平滑。重点抽样模拟量为

$$\mathrm{E}[f(\boldsymbol{\varphi})] \approx \frac{1}{R}\sum_{r=1}^{R}\frac{h(\boldsymbol{\varphi}^{(r)})f(\boldsymbol{\varphi}^{(r)})}{g(\boldsymbol{\varphi}^{(r)})}$$

其中,$\boldsymbol{\varphi}^{(r)}$ 为从重点密度 $g(\boldsymbol{\varphi})$ 中的一次抽取。

考虑一个随机截距模型，重点取样器构造为

$$\int_{-\infty}^{\infty} \phi(v_j;\mu_i,\tau_j^2)\left[\frac{\phi(v_j)\prod g^{(1)}(y_{ij}\mid v_j)}{\phi(v_j;\mu_i,\tau_j^2)}\right]dv_j$$

其中，重点密度 $\phi(v_j;\mu_i,\tau_j^2)$ 为近似后验密度的正态密度。样本从 $\phi(v_j;\mu_j,\tau_j^2)$ 中抽取以计算方括号中项的均值。注意，这种方法类似于自适应积分，正如皮涅罗和贝茨(Pinheiro and Bates,1995)所指出，自适应积分可被视为重点抽样的确定形式。重点取样器的多元扩展由库克(Kuk,1999)和司考戈(Skaug,2002)用于广义线性混和模型。

有趣的是，蒙特卡洛积分是**准蒙特卡洛**(quasi-Monte Carlo)，其样本可确定性抽取，详见 Shaw(1988)以及 Fang and Wang(1994,第2章)。

## 6.3.4 专用模拟器：GHK

一些模拟器专用于特殊的模型，如适用于诸如多项正态概率模型和多元正态概率模型等多元正态潜响应模型的 Geweke-Hajivassiliou-Keane (GHK) (Geweke,1989;Hajivassiliou and Ruud,1994;Keane,1994) 以及斯特恩模拟器(Stern,1992)。

对于这些模型，似然贡献为如下形式的概率，这里用 $p$ 表示，即

$$p = \Pr[(\tau_1^- < \epsilon_1 \leqslant \tau_1^+),(\tau_2^- < \epsilon_2 \leqslant \tau_2^+),\cdots,(\tau_S^- < \epsilon_S \leqslant \tau_S^+)]$$

其中，$\boldsymbol{\epsilon}=(\epsilon_1,\epsilon_2,\cdots,\epsilon_S)'$是均值为零，方差矩阵为 $\boldsymbol{\Sigma}$ 的 $S$ 维多元正态向量。在**多元**正态概率情形下，$\boldsymbol{\epsilon}$ 代表潜响应 $\boldsymbol{y}^*$ 的残差，并且在**多项**正态概率的情形下，表示效用残差的差异。积分上下限或阈值 $\tau_s$ 特别取决于协变量。例如，在针对定序响应或二项响应的多元正态概率模型中，$\tau_s=\kappa_s-\mathrm{x}'\boldsymbol{\beta}$，其中，$\kappa_s$ 为阈值模型参数(见2.4节)。

注意概率 $p$ 是对潜响应分布的矩形域积分，不是对潜变量分布积分。相反，本节中描述的其他方法都采用潜变量分布积分(见4.9节对于潜响应和潜变量积分的讨论)。潜响应积分也用于将在6.7节中讨论的有限信息方法中。

这里我们主要介绍通用的 GHK 模拟量。首先，我们认为概率 $p$ 可用一系列条件一元正态分布乘积来表达，定义为

$$Q_1 \equiv \Pr[(\tau_1^- < \epsilon_1 \leqslant \tau_1^+)]$$

$$Q_2 \equiv \Pr[(\tau_2^- < \epsilon_2 \leqslant \tau_2^+)]\mid[(\tau_1^- < \epsilon_1 \leqslant \tau_1^+)]$$

$$\vdots$$

$$Q_s \equiv \Pr[(\tau_S^- < \epsilon_S \leqslant \tau_S^+)\mid(\tau_{S-1}^- < \epsilon_{S-1} \leqslant \tau_{S-1}^+),\cdots,(\tau_1^- < \epsilon_1 \leqslant \tau_1^+)]$$

概率可写为

$$p = \prod_{s=1}^{S} Q_s$$

很容易计算出 $Q_1=\Phi(\tau_1^+/\sigma_{11})-\Phi(\tau_1^-/\sigma_{11})$，其中，$\Phi(\cdot)$表示一元累积标准正态分布函数且 $\sigma_{11}$是 $\epsilon_1$ 的标准差。然而，给定其他在特定区间内的 $\epsilon_t$(与 $\epsilon_s$ 相关)，对于每个 $Q_s, s=2,\cdots S$，都是 $\epsilon_s$ 处在一个区间内的**条件**概率，这是难于计算的。

因此，我们对残差 $\epsilon$ 正交化——采用协方差矩阵 $\boldsymbol{\Sigma}$ 的低阶对角乔莱斯基分解，$\boldsymbol{\Sigma}=\boldsymbol{CC}'$，其元素为 $c_{sm}$，当 $m>s$ 时，$c_{sm}=0$。当 $u$ 具有**独立**标准正态分量 $u_s$ 时，我们可写为$\epsilon=Cu$，其中 $u$ 为正交向量。

算法流程如下：

(1)对于迭代 $r=1,\cdots,R$：

(a)对 $s=1$；

- 估计 $Q_{1r}$：

$$Q_{1r} = \Phi(\tau_1^+/c_{11}) - \Phi(\tau_1^-/c_{11})$$

• 由具有截断点 $\tau_1^-/c_{11}$ 和 $\tau_1^+/c_{11}$ 的**双截断**标准正态分布模拟 $u_{1r}$，因此，对条件概率 $Q_2$ 到 $Q_s$，$\epsilon_{1r}=c_{11}u_{1r}$ 满足条件 $\tau_1^- < \epsilon_{1r} \leq \tau_1^+$。

(b)对 $s=2$；

• 估计 $Q_{2r}$

将 $u_{1r}$ 作为已知，$Q_{2r}$ 则变为易估计的**非条件**概率

$$Q_{2r} = \Phi([\tau_2^+ - c_{21}u_{1r}]/c_{22}) - \Phi([\tau_2^- - c_{21}u_{1r}]/c_{22})$$

第一项遵循如下等价

$$\epsilon_{2r} = c_{21}u_{1r} + c_{22}u_{2r} \leq \tau_2^+ \Leftrightarrow u_{2r} \leq [\tau_2^+ - c_{21}u_{1r}]/c_{22}$$

第二项也如此。

• 由具有截断点 $[\tau_2^- - c_{21}u_{1r}]/c_{22}$ 和 $[\tau_2^+ - c_{21}u_{1r}]/c_{22}$ 的双截断标准正态分布模拟 $u_{2r}$，因此对条件概率 $Q_3$ 到 $Q_s$，$\epsilon_{2r}$ 满足条件。

(c)对 $s = 3,\cdots,S$：

继续估计 $Q_{sr}$，将 $u_{1r}$ 至 $u_{s-1,r}$ 作为已知：

$$Q_{sr} = \Phi([\tau_s^+ - \sum_{m=1}^{s-1} c_{sm}u_{mr}]/c_{ss}) - \Phi([\tau_s^- - \sum_{m=1}^{s-1} c_{sm}u_{mr}]/c_{ss})$$

第一项遵循如下等效

$$\epsilon_{sr} = \sum_{m=1}^{s} c_{sm}u_{mr} \leq \tau_s^+ \Leftrightarrow u_{sr} \leq [\tau_s^+ - \sum_{m=1}^{s-1} c_{sm}u_{mr}]/c_{ss}$$

• 由具有截断点 $[\tau_s^- - \sum_{m=1}^{s-1} c_{sm}u_{mr}]/c_{ss}$ 和 $[\tau_s^+ - \sum_{m=1}^{s-1} c_{sm}u_{mr}]/c_{ss}$ 的双截断标准正态分布模拟 $u_{sr}$，因此，对条件概率 $Q_{s+1}$ 到 $Q_s$，$\epsilon_{sr}$ 满足条件(当 $s=S$ 时，此步骤不需要)。

(2)在 R 次重复后：所需的模拟概率可由下式得到

$$\check{p} = \frac{1}{R}\sum_{r=1}^{R}\prod_{s=1}^{S} Q_{sr}$$

注意，当期望超过假定的模拟重复次数时，$\check{p}$ 是具有 $E(\check{p})=p$ 的 $p$ 的无偏模拟量。然而，模拟的对数似然是具有 $\log(\check{p})$ 形式的项的和，并且它们通常是有偏的，因为非线性对数变换，$E[\log(\check{p})] \neq \log(p)$。偏差随着重复次数 $R$ 的增加而减少。

GHK 模拟量通常和梯度方法协同使用(将在 6.4.2 节中讨论)以获得**最大模拟似然**(maximum simulated likelihood, MSL)估计量。关于 MSL 的性质及相关估计量的进一步讨论请参考 Train (2003)。

在计量经济学中，GHK 模拟量在多元正态潜响应模型中使用非常广泛，如**正态概率面板(纵贯性)**模型(如 Keane, 1994; Geweke et al., 1994)。这可能是因为该模拟量尽管相对容易实现，却依然胜过其他模拟量。与勒曼和曼斯基(Lerman and Manski, 1981)最早提出的原始蒙特卡洛方法不同，GHK 模拟量是连续且可微参数函数，并且可得出无偏且界限处于(0,1)区间中的模拟概率。此外，GHK 比其他模拟量(如 Hajivassiliou et al., 1996)更加统计有效。关于 GHK 以及相关模拟量的非技术讨论，请参考 Stern (1997)、Train (2003)以及 Cappellari and Jenkins (2003)。

## 6.4 似然最大化

有很多方法可以对似然最大化，最普遍的方法是期望最大化算法(EM)和牛顿-拉夫逊算法或者费舍尔计分算法，将在 6.4.2 节和 6.4.1 节中介绍。上述每种积分方法将伴随最大化方法一同介绍。

### 6.4.1 EM 算法

期望最大化(EM)算法是一种对不完备数据问题求最大似然估计的迭代算法。此算法由丹普斯特等人(Dempster et al.,1977)命名,他们提出了这种算法的广义理论和一系列例子。奥奇德和伍德贝里(Orchard and Woodbury,1972)首次注意到所谓的"缺失信息法则"这一潜在想法的广泛适用性,尽管 EM 算法的应用可追溯到麦肯迪克(McKendrick,1926)。

也许突出的应用在于当实现需格外注意的随机变量缺失数据时的估计(如 Little and Rubin,2002)。另一个在现有框架下更为重要的应用是潜变量模型的估计。在这种情况下,潜变量的实现被理解为缺失数据(如 Becker et al.,1997)。

EM 算法背后诱发的想法如下:观测数据随能允许用一系列简单估计步骤来估计的潜数据而增加,而非通过一次复杂的估计。

完整数据 $\mathbf{C}=\{\boldsymbol{y},\mathbf{X},\boldsymbol{\zeta}\}$ 由两部分组成:可观测的不完整数据 $\boldsymbol{y}$ 和 $\mathbf{X}$ 以及不可观测到的数据或潜数据 $\boldsymbol{\zeta}$。假定潜数据是可观测的,则完整数据对数似然用 $l_h(\boldsymbol{\vartheta}|\mathbf{C})$ 表示。由于对数似然是李和奈尔德(Lee and Nelder,1996)提出的 $h$-对数似然,因此,这里我们使用 $h$ 下标。总之,完整数据对数似然本身包含于 EM 算法的每一步积分中,在第$(k+1)$步有如下形式:

**E 步进**:计算后验期望

$$Q(\boldsymbol{\vartheta}\mid\boldsymbol{\vartheta}^k)\equiv \mathrm{E}_{\zeta}[l_h(\boldsymbol{\vartheta}\mid\mathbf{C})\mid\boldsymbol{y},\mathbf{x};\boldsymbol{\vartheta}^k]$$

给定非完整数据及由前一次迭代所得估计 $\boldsymbol{\vartheta}^k$,关于潜变量的完整数据对数似然的条件期望,如关于 $\boldsymbol{\zeta}$ 后验密度的期望。

**M 步进**:关于 $\boldsymbol{\vartheta}$ 最大化 $Q(\boldsymbol{\vartheta}|\boldsymbol{\vartheta}^k)$,从而得到新的估计 $\boldsymbol{\vartheta}^{k+1}$。

这有时采用解析完成,但通常会使用诸如梯度方法(见 6.4.2 节)的迭代算法。

现在我们来考虑具有多元正态潜变量的双层次潜变量模型的 EM 算法的实现。正如 6.3.2 节所说的那样,通过乔莱斯基分解 $\boldsymbol{\zeta}_j=\mathbf{Q}\mathbf{v}_j$(其中 $\mathbf{Q}$ 取决于 $\boldsymbol{\vartheta}$)变积分变量为独立标准正态潜变量 $\mathbf{v}_j=(v_{j1},v_{j2},\cdots,v_{jM})'$是非常方便的。将正交化的潜变量 $\mathbf{v}_j$ 当做观测变量,则完整数据对数似然可表示为

$$\begin{aligned}
l_h(\boldsymbol{\vartheta}\mid\mathbf{C}) &= \ln\prod_j\left\{\prod_i g(y_{ij}\mid\mathbf{Q}\mathbf{v}_j)\prod_m\phi(v_{jm})\right\}\\
&= \sum_j\left\{\sum_i \ln g(y_{ij}\mid\mathbf{Q}\mathbf{v}_j)+\sum_m\ln\phi(v_{jm})\right\} \qquad (6.13)\\
&= \sum_j l_h^j(\boldsymbol{\vartheta}\mid\mathbf{C})
\end{aligned}$$

其中,$l_h^j(\boldsymbol{\vartheta}|\mathbf{C})$为对于完整数据对数似然的群组贡献。

**E 步进**:计算

$$\begin{aligned}
Q(\boldsymbol{\vartheta}\mid\boldsymbol{\vartheta}^k) &= \mathrm{E}_{\zeta}[l_h(\boldsymbol{\vartheta}\mid\mathbf{C})\mid\mathbf{y};\boldsymbol{\vartheta}^k]\\
&= \sum_j\int l_h^j(\boldsymbol{\vartheta}\mid\mathbf{C})\omega(\mathbf{v}_j\mid\mathbf{y}_{j(2)};\boldsymbol{\vartheta}^k)\mathrm{d}\mathbf{v}_j
\end{aligned}$$

其中,$\omega(\mathbf{v}_j|\mathbf{y}_{j(2)};\boldsymbol{\vartheta}^k)$为在给定群组的观测响应 $\mathbf{y}_{j(2)}$ 条件下,潜变量 $\mathbf{v}_j$ 关于群组 $j$ 的后验密度。使用贝叶斯定理,后验密度就为

$$\omega(\mathbf{v}_j\mid\mathbf{y}_{j(2)};\boldsymbol{\vartheta}^k)=\frac{\prod_i g(y_{ij}\mid\mathbf{Q}^k\mathbf{v}_j;\boldsymbol{\vartheta}^k)\prod_m\phi(v_{jm})}{\int\prod_i g(y_{ij}\mid\mathbf{Q}^k\mathbf{v}_j;\boldsymbol{\vartheta}^k)\prod_m\phi(v_{jm})\mathrm{d}\mathbf{v}_j} \qquad (6.14)$$

其中，$\mathbf{Q}^k$ 中的 $k$ 上标表示该矩阵依赖于 $\boldsymbol{\vartheta}^k$。使用式(6.13)和式(6.14)，$Q(\boldsymbol{\vartheta}|\boldsymbol{\vartheta}^k)$简化为

$$Q(\boldsymbol{\vartheta}|\boldsymbol{\vartheta}^k)=\sum_j\frac{1}{A_j^k}\int_{-\infty}^{\infty}\left\{\sum_i\ln g(y_{ij}|\mathbf{Q}\mathbf{v}_j;\boldsymbol{\vartheta})+\sum_m\ln\phi(v_{jm})\right\}\times$$
$$\prod_i g(y_{ij}|\mathbf{Q}^k\mathbf{v}_j;\boldsymbol{\vartheta}^k)\prod_m\phi(v_{jm})\,\mathrm{d}\mathbf{v}_j$$

其中

$$A_j^k=\int_{-\infty}^{\infty}\prod_i g(y_{ij}|\mathbf{Q}^k\mathbf{v}_j;\boldsymbol{\vartheta}^k)\prod_m\phi(\mathrm{v}_{jm})\,\mathrm{d}\mathbf{v}_j$$

注意 $A_j^k$ 并不依赖于未知参数 $\boldsymbol{\vartheta}$，而是仅仅依赖于在前一次迭代中获得的 $\boldsymbol{\vartheta}^k$。

通常当积分不能获得解析解时，E 步进会变得复杂，此时建议使用一些基于数值积分的近似方法或模拟。可用蒙特卡洛积分（如 Wei and Tanner，1990）来得到蒙特卡洛期望最大化（Monte Carlo Expectation Maximization，MCEM）算法。在本文中，对于独立正态分布随机变量 $\mathbf{v}_j$，其复制 $\mathrm{r}=1,2,\cdots,R$，我们考虑抽取向量 $\mathbf{d}_{jr}=(\mathrm{d}_{jr1},\mathrm{d}_{jr2},\cdots,\mathrm{d}_{jrM})'$。它可提供如下蒙特卡洛积分近似

$$Q^{MC}(\boldsymbol{\vartheta}|\boldsymbol{\vartheta}^k)=\sum_j\sum_r c_{jr}^{MC}\left\{\sum_i\ln g(y_{ij}|\mathbf{Q}\mathbf{d}_{jr};\boldsymbol{\vartheta})+\sum_m\ln\phi(d_{jrm})\right\}$$

其权重不依赖于未知参数 $\boldsymbol{\vartheta}$，且

$$c_{jr}^{MC}=\frac{\sum_i g(y_{ij}|\mathbf{Q}^k\mathbf{d}_{jr};\boldsymbol{\vartheta}^k)}{\sum_r\sum_i g(y_{ij}|\mathbf{Q}^k\mathbf{d}_{jr};\boldsymbol{\vartheta}^k)}$$

满足 $\sum_r c_{jr}^{MC}=1$。除了使用原始蒙特卡洛积分以外，孟和希林（Meng and Schilling，1996）提出使用吉布斯抽样（Gibbs sampling）（见 6.11.5 节）。MCEM 的一个缺陷在于无法量化由算法的每一步所带来的蒙特卡洛误差（如 McCulloch，1997；Hobert，2000）。如果重复次数 $R$ 太小，则 E 步进会陷入蒙特卡洛误差，反之过大的 $R$ 则会造成浪费。事实上，布斯等人（Booth et al.，2001）指出，在大多数情形下，除非 $R$ 随着 $k$ 增加，否则 MCEM 不会收敛。

既然潜变量分布被设定为正态的，那么也可使用高斯-埃尔米特积分，从而得到

$$Q^{GH}(\boldsymbol{\vartheta}|\boldsymbol{\vartheta}^k)=\sum_j\sum_{r_1}c_{r_1}^{GH}\cdots\sum_{r_M}c_{r_M}^{GH}\left\{\sum_i\ln g(y_{ij}|\mathbf{Q}\mathbf{a}_r;\boldsymbol{\vartheta})+\sum_m\ln\phi(a_{rm})\right\}$$

其权重不依赖于 $\boldsymbol{\vartheta}$，即

$$c_{r_m}^{GH}=\frac{p_{r_m}\sum_i g(y_{ij}|\mathbf{Q}^k\mathbf{a}_r;\boldsymbol{\vartheta}^k)}{\sum_r p_{r_m}\sum_i g(y_{ij}|\mathbf{Q}^k\mathbf{a}_r;\boldsymbol{\vartheta}^k)}$$

满足 $\sum_{r_m}c_{r_m}^{CH}=1$。在此，$\mathbf{a}_r=(a_{r1},a_{r2},\cdots,a_{rM})'$和 $p_{r_m}=p_r$ 分别代表积分位置和权重。博科和希林（Bock and Schilling，1997）建议使用自适应积分来改进积分近似。

**M 步进**：对 $\boldsymbol{\vartheta}$ 最大化 $Q(\boldsymbol{\vartheta}|\boldsymbol{\vartheta}^k)$。

对 MCEM，则等价于求解如下方程

$$\frac{\partial Q^{MC}(\boldsymbol{\vartheta}|\boldsymbol{\vartheta}^k)}{\partial\boldsymbol{\vartheta}}=\sum_j\sum_r c_{jr}^{MC}\frac{\partial\sum_i\ln g(y_{ij}|\mathbf{Q}\mathbf{d}_{jr};\boldsymbol{\vartheta})}{\partial\boldsymbol{\vartheta}}=\mathbf{0}$$

这是一个在针对扩展数据 $y_{ij}$的广义线性模型下，关于 $\boldsymbol{\vartheta}$ 的加权得分函数 $d_{jr}$，且有已知权重 $c_{jr}^{MC}$，则 $Q(\boldsymbol{\vartheta}|\boldsymbol{\vartheta}^k)$可简单地用标准软件通过加权最大似然得到最大化。

对于高斯-埃尔米特积分，M 步进总计求解

$$\frac{\partial Q^{GH}(\boldsymbol{\vartheta} \mid \boldsymbol{\vartheta}^k)}{\partial \boldsymbol{\vartheta}} = \sum_j \sum_{r_1} c_{r_1}^{GH} \cdots \sum_{r_M} c_{r_M}^{GH} \frac{\partial \sum_i \ln g(y_{ij} \mid \mathbf{Qa}_r; \boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} = \mathbf{0}$$

这里也可用适当的加权最大似然算法最大化。请见艾特金(Aitkin,1999a)针对双层次模型对这个算法的实现,以及沃芒(Vermunt,2004)在更多层次模型上的算法。

通常,E 步进较为费力,但在一些情况下,M 步进则更加困难。一种简化方法为 ECM 算法(Meng and Rubin,1993),将每一个 M 步进替换为一系列条件最大化步骤,并且 $\boldsymbol{\vartheta}$ 的子集被固定为其在前一步中的值。雷托和鲁宾(Little and Rubin,2002)讨论了一些其他的修正方法。另一种可能是在 M 步进中使用模拟,即使用原始蒙特卡洛积分或通过更加有针对性的算法,如麦克拉奇(McCulloch 1994,1997)提出的 Metropolis 算法(见 6.11.5 节)或布斯和霍伯特(Booth and Hobert, 1999)提出的重点抽样等更精细的方法。

值得注意的是,当完整数据分布呈**正则指数族**形式时,EM 效果最好。在这种情况下可证明(如 Tanner,1996),E 步进包括由后验期望估计的完整数据充分统计量。给定这些估计,M 步进的似然方程则和完整数据的似然方程有相同的形式,因此可使用标准软件。

在 EM 算法的实现极其简单的情况下,传统探索性因子模型的估计是值得考虑的。用闭式表达的期望的充分统计量可通过 E 步进得到,而在 M 步进中只需要基本的线性代数。

**例:EM 探索性因子分析**

考虑 3.3.3 节中介绍的探索性因子模型

$$\mathbf{y}_j = \boldsymbol{\Lambda}\boldsymbol{\eta}_j + \boldsymbol{\epsilon}_j$$

其中,$\mathbf{y}_j$ 为均值中心化的响应,$\boldsymbol{\eta}_j \sim \mathrm{N}_M(\mathbf{0},\mathbf{I})$,$\boldsymbol{\epsilon}_j \sim \mathrm{N}_n(\mathbf{0},\boldsymbol{\Theta})$,$\boldsymbol{\eta}_j$ 与 $\boldsymbol{\epsilon}_j$ 是独立的,$\boldsymbol{\Theta}$ 是对角矩阵,$\boldsymbol{\Lambda}$ 是无结构的。

由探索性因子模型可得

$$\mathbf{y}_j \sim \mathrm{N}_n(\mathbf{0},\boldsymbol{\Sigma})$$

其中

$$\boldsymbol{\Sigma} \equiv \mathrm{Cov}(\mathbf{y}_j) = \boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}$$

另外,我们可得到

$$\mathbf{y}_j \mid \boldsymbol{\eta}_j \sim \mathrm{N}_n(\boldsymbol{\Lambda}\boldsymbol{\eta}_j, \boldsymbol{\Theta})$$

重要的是,如果公共因子是可观测的,则探索性因子模型则变为一个关于响应 $\mathbf{y}_j$,标准协变量 $\boldsymbol{\eta}_j$,回归参数 $\boldsymbol{\Lambda}$,残差协方差矩阵 $\boldsymbol{\Theta}$ 的一个多元回归模型。完整数据对数似然因此等同于具有下述充分统计量的正态多元回归,即

$$S_{yy} = \frac{1}{N}\sum_{j=1}^{N} \mathbf{y}_j\mathbf{y}_j'$$

$$S_{y\eta} = \frac{1}{N}\sum_{j=1}^{N} \mathbf{y}_j\boldsymbol{\eta}_j'$$

以及

$$S_{\eta\eta} = \frac{1}{N}\sum_{j=1}^{N} \boldsymbol{\eta}_j\boldsymbol{\eta}_j'$$

给定响应,这些因子的条件分布变成

$$\boldsymbol{\eta}_j \mid \mathbf{y}_j \sim \mathrm{N}_M(\widetilde{\boldsymbol{\eta}}_j, \mathbf{Y})$$

其中

$$\tilde{\boldsymbol{\eta}}_j \equiv \mathrm{E}(\boldsymbol{\eta}_j \mid \mathbf{y}_j) = \mathbf{F}\mathbf{y}_j$$

$$\Upsilon \equiv \mathrm{Cov}(\boldsymbol{\eta}_j \mid \mathbf{y}_j) = \mathbf{I} - \mathbf{F}\boldsymbol{\Lambda}$$

**F** 为因子计分的回归方法的“因子计分矩阵”(见 7.3 节),即

$$\mathbf{F} \equiv \boldsymbol{\Lambda}'\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Lambda}'(\boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{\Theta})^{-1}$$

**E** 步进:给定响应 **y** 和参数 $\boldsymbol{\vartheta}^k$,充分统计量的条件期望(关于因子 $\eta_j$)则变为

$$\mathrm{E}_\eta(\mathbf{S}_{yy} \mid \mathbf{y}, \boldsymbol{\vartheta}^k) = \frac{1}{N}\sum_{j=1}^{N}\mathrm{E}_\eta(\mathbf{y}_j\mathbf{y}_j' \mid \mathbf{y}, \boldsymbol{\vartheta}^k) = \mathbf{S}_{yy}$$

$$\mathrm{E}_\eta(\mathbf{S}_{y\eta} \mid \mathbf{y}, \boldsymbol{\vartheta}^k) = \frac{1}{N}\sum_{j=1}^{N}\mathrm{E}_\eta(\mathbf{y}_j\eta_j' \mid \mathbf{y}, \boldsymbol{\vartheta}^k) = \mathbf{S}_{yy}\mathbf{F}^{k'}$$

且

$$\mathrm{E}_\eta(\mathbf{S}_{\eta\eta} \mid \mathbf{y}, \boldsymbol{\vartheta}^k) = \frac{1}{N}\sum_{j=1}^{N}\mathrm{E}_\eta(\eta_j\eta_j' \mid \mathbf{y}, \boldsymbol{\vartheta}^k) = \mathbf{F}^k\mathbf{S}_{yy}\mathbf{F}^{k'} + \boldsymbol{\Upsilon}^k$$

其中

$$\mathbf{F}^k \equiv \boldsymbol{\Lambda}^{k\prime}(\boldsymbol{\Lambda}^k\boldsymbol{\Lambda}^{k\prime} + \boldsymbol{\Theta}^k)^{-1}$$

以及

$$\Upsilon^k \equiv \mathbf{I} - \mathbf{F}^k\boldsymbol{\Lambda}^k$$

**M** 步进:将标准多元回归估计量按照从 E 步进(取代通常的充分统计量)得到的期望充分统计量展开:

$$\begin{aligned}
\boldsymbol{\Lambda}^{k+1} &= \mathrm{E}_\eta(\mathbf{S}_{y\eta} \mid \mathbf{y}, \boldsymbol{\vartheta}^k)\{\mathrm{E}_\eta(\mathbf{S}_{\eta\eta} \mid \mathbf{y}, \vartheta^k)\}^{-1} \\
&= \mathbf{S}_{yy}\mathbf{F}^{k\prime}(\mathbf{F}^k\mathbf{S}_{yy}\mathbf{F}^{k\prime} + \Upsilon^k)^{-1} \\
\boldsymbol{\Theta}^{k+1} &= \mathrm{diag}\Big(\mathrm{E}_\eta(\mathbf{S}_{y\eta} \mid \mathbf{y}, \boldsymbol{\vartheta}^k) - \mathrm{E}_\eta(\mathbf{S}_{y\eta} \mid \mathbf{y}, \boldsymbol{\vartheta}^k) \\
&\qquad [\mathrm{E}_\eta(\mathbf{S}_{\eta\eta} \mid \mathbf{y}, \boldsymbol{\vartheta}^k)]^{-1}\mathrm{E}_\eta(\mathbf{S}_{y\eta} \mid \mathbf{y}, \boldsymbol{\vartheta}^k)'\Big) \\
&= \mathrm{diag}(\mathbf{S}_{yy} - \mathbf{S}_{yy}\mathbf{F}^{k\prime}(\mathbf{F}^k\mathbf{S}_{yy}\mathbf{F}^{k\prime} + \boldsymbol{\Upsilon}^k)^{-1}\mathbf{F}^k\mathbf{S}_{yy}) \\
&= \mathrm{diag}(\mathbf{S}_{yy} - \boldsymbol{\Lambda}^{k+1}\mathbf{F}^k\mathbf{S}_{yy})
\end{aligned}$$

这种使用 EM 算法求因子模型估计的思想缘自丹普斯特等人(Dempster et al., 1977)。鲁宾和塞亚尔(Rubin and Thayer, 1982)以及肖恩伯格和里奇谭德(Schoenberg and Richtand, 1984)做了详细的介绍并扩展到验证性因子模型(见 3.3.3 节)。刘等人(Liu et al., 1998)证明了上述方法是所谓参数扩展 EM(PXEM)算法的一个特例。陈(Chen, 1981)介绍了一种用 EM 算法的传统 MIMIC 模型的估计。

**续例:EM 探索性因子分析**

EM 算法也可用另一种更符合直觉的但有些不同的方式来实现。注意,上述关于 $\boldsymbol{\Lambda}^{k+1}$ 和 $\boldsymbol{\Theta}^{k+1}$ 的方程与 $\mathbf{y}_j$ 对 $\tilde{\eta}_j^k = \mathbf{F}^k\mathbf{y}_j$ 的线性回归相关,我们可用下面的迭代算法:

**E** 步进:用 $\tilde{\eta}_j^k$ 推算缺失的 $\eta_j$。这也是将在 7.3.1 节中介绍的用于因子计分的“回归方法”的一个例子。

**M** 步进:通过 $y_{ij}$($\mathbf{y}_j$ 的第 $i$ 项)对 $\tilde{\eta}_j^k$ 的 OLS 回归,估计 $\boldsymbol{\Lambda}_i^{k+1}$($\boldsymbol{\Lambda}$ 的第 $i$ 行)以及 $\theta_{ii}^{k+1}$。这种简化缘自模型有一个对角残差协方差矩阵 $\boldsymbol{\Theta}$。

该方程阐明了 EM 算法可视为一种处理缺失数据的直觉方法的形式化:①用预测值推算缺失值;②将推算值作为数据来估计参数;③将新估计作为真实参数重新推算缺失值;④重新估计参数,直到收敛。谨记这种基于推算的方法仅适用于当完整数据似然方

程在缺失数据中是线性的情况下。否则，此方法会造成严格有偏估计。

在指数族之外，E 步进的数值误差在 M 步进中，对于待被最大化的函数易产生人为众数（如，Meng and Rubin，1992）。由于这个缘故，孟和希林（Meng and Schilling，1996）使用数据增强算法，生成潜响应 $\mathbf{y}^*$ 以修正博科和艾特金（Bock and Aitkin，1981）的正态概率因子模型中的 E 步进。尽管对于 $\mathbf{y}$ 的模型并不是指数族的一员，但是孟和希林利用了潜响应 $\mathbf{y}^*$ 的完整数据分布属于指数族形式这一特点。另见 6.11.5 节关于正态概率模型中数据增强算法的一个例子。

EM 算法被广泛应用于潜变量模型中。例如，丹普斯特等人（Dempster et al.，1981）和陈（Chen，1981）分别提出的针对连续响应的因子模型和 MIMIC 模型，Bock and Aitkin（1981）的探索性正态概率因子模型，Thissen（1982）的单参数对数概率 IRT 模型。Strenio et al.（1983）以及 Raudenbush and Bryk（2002）研究了线性混和模型，Aitkin et al.（1981），Aitkin（1999a）以及 Vermunt（2004）研究了广义线性混和模型。EM 算法是估计潜类模型和有限混和模型（如 Goodman，1974）的最常用的方法。

EM 算法通常被提及的优点在于较之其他优化方法更灵活的实现性。尽管这在很多情况下是毫无疑问的，但从上述推导的公式可显而易见地得出，这种结论在复杂的潜变量模型中并不明显。它的理论优势还包括每一个迭代都增大似然，并且如果序列 $\boldsymbol{\vartheta}^k$ 收敛，则会收敛于局部最大或鞍点。

EM 算法一个重要的缺点在于，当存在较大比例的缺失信息时收敛速度非常慢。另一个缺点在于较之牛顿-拉夫逊之类的梯度方法，EM 算法所估计的信息矩阵不是最大化的直接副产品。一个可行的方法是在 EM 算法收敛后增加最后一个牛顿-拉夫逊步进。路易斯（Louis，1982）、孟和鲁宾（Meng and Rubin，1991）以及奥克斯（Oakes，1999）等提供了在 EM 算法中获取信息矩阵的方法。

## 6.4.2 梯度方法

为使对数似然最大化，我们必须求解似然方程

$$\frac{\partial l(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} = \mathbf{0}$$

梯度方法是迭代式的，我们将第 $k$ 次迭代中的参数用 $\boldsymbol{\vartheta}^k$ 表示。

**牛顿-拉夫逊算法和费舍尔计分**

用当前参数估计 $\boldsymbol{\vartheta}^k$ 处的一阶泰勒级数展开作为对数似然导数的近似，可导出**牛顿-拉夫逊算法**：

$$\begin{aligned}\frac{\partial l(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} &\approx \frac{\partial l(\boldsymbol{\vartheta}^k)}{\partial \boldsymbol{\vartheta}} + \frac{\partial^2 l(\boldsymbol{\vartheta}^k)}{\partial \boldsymbol{\vartheta}\partial \boldsymbol{\vartheta}'}(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}^k)\\ &= \mathbf{g}(\boldsymbol{\vartheta}^k) + \mathbf{H}(\boldsymbol{\vartheta}^k)(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}^k)\end{aligned} \tag{6.15}$$

其中，$\mathbf{g}(\boldsymbol{\vartheta}^k)$ 为 $V$-维梯度向量，$\mathbf{H}(\boldsymbol{\vartheta}^k)$ 为海森矩阵，即对数似然关于估计参数 $\boldsymbol{\vartheta}^k$ 的二阶导数 $v \times v$ 矩阵。校正参数 $\boldsymbol{\vartheta}^{k+1}$ 是使一阶泰勒展开为零的参数，如

$$\mathbf{g}(\boldsymbol{\vartheta}^k) + \mathbf{H}(\boldsymbol{\vartheta}^k)(\boldsymbol{\vartheta}^{k+1} - \boldsymbol{\vartheta}^k) = \mathbf{0}$$

因此

$$\boldsymbol{\vartheta}^{k+1} = \boldsymbol{\vartheta}^k - \mathbf{H}(\boldsymbol{\vartheta}^k)^{-1}\mathbf{g}(\boldsymbol{\vartheta}^k)$$

注意，若对数似然是这些参数的二次式，在这种情况下，最大值可用一次迭代找到，故而泰勒展开是精确的。典型的例子如标准线性回归模型。

**费舍尔计分算法**类似于牛顿-拉夫逊算法，不过用费舍尔信息矩阵 $\mathrm{I}(\boldsymbol{\vartheta}^k)$ 的负矩阵来

取代海森矩阵,如

$$\boldsymbol{\vartheta}^{k+1} = \boldsymbol{\vartheta}^{k} + \mathbf{I}(\boldsymbol{\vartheta}^{k})^{-1}\mathbf{g}(\boldsymbol{\vartheta}^{k})$$

其中

$$\mathbf{I}(\boldsymbol{\vartheta}^{k}) = -\mathrm{E}(\mathbf{H}(\boldsymbol{\vartheta}^{k}))$$

与 EM 算法相比,牛顿-拉夫逊算法和费舍尔计分算法的优点在于他们可提供最大似然估计 $\hat{\boldsymbol{\vartheta}}$ 的标准误的估计。在费舍尔计分算法中,使用逆信息矩阵;而在牛顿-拉夫逊算法中,使用 $-\mathbf{H}(\hat{\boldsymbol{\vartheta}})$ 的逆矩阵,即"观测信息"(另见 8.3 节)。

**准牛顿方法**

牛顿-拉夫逊算法和费舍尔计分算法需要对数似然关于参数的二阶导数。解析计算这些是非常困难的并且数值计算非常缓慢。因此,提出了只需要梯度的**准牛顿方法**。

博恩特、霍尔、霍尔与豪斯曼(Berndt, Hall, Hall and Hausman, 1974)提出了一种有用的算法。他们提出的 BHHH 或 $\mathrm{BH}^3$ 算法基于这个事实,即在正确的模型设定下,信息矩阵等于梯度的协方差矩阵,则

$$\mathbf{I}(\boldsymbol{\vartheta}^{k}) = -\mathrm{E}(\mathbf{H}(\boldsymbol{\vartheta}^{k})) = \mathrm{E}(\mathbf{g}(\boldsymbol{\vartheta}^{k})\mathbf{g}(\boldsymbol{\vartheta}^{k})')$$

由大数定理可得,协方差矩阵的一致估计量为

$$\mathbf{I}_{\mathrm{BH}^3}(\boldsymbol{\vartheta}^{k}) \equiv \sum_{z=1}^{J}\mathbf{g}^{z}(\boldsymbol{\vartheta}^{k})\mathbf{g}^{z}(\boldsymbol{\vartheta}^{k})'$$

其中,$\mathbf{g}^{z}(\boldsymbol{\vartheta}^{k})$ 为得分向量,顶层群组对于梯度的贡献为

$$\mathbf{g}(\boldsymbol{\vartheta}^{k}) = \sum_{z=1}^{Z}\mathbf{g}^{z}(\boldsymbol{\vartheta}^{k})$$

$\mathrm{BH}^3$ 算法使用费舍尔计分算法中的估计量,即

$$\boldsymbol{\vartheta}^{k+1} = \boldsymbol{\vartheta}^{k} + [\mathbf{I}_{\mathrm{BH}^3}(\boldsymbol{\vartheta}^{k})]^{-1}\mathbf{g}(\boldsymbol{\vartheta}^{k})$$

$\mathrm{BH}^3$ 算法的优点在于只要求梯度;不需要计算海森矩阵和费舍尔信息矩阵。雷托和鲁宾(Little and Rubin, 2002)指出 $\mathrm{BH}^3$ 算法的性能会不稳定,因为信息矩阵近似的精度取决于模型的有效性,但我们的经验表明,即使使用了非常"糟糕"的初始值,这个算法也工作良好。

其他使用准牛顿算法的例子有 Davidon-Fletcher-Powell(DFP)算法,以及对海森矩阵 $\mathbf{H}(\boldsymbol{\vartheta}^{k})$ 使用不同近似的 Broyden-Fletcher-Goldfarb-Shanno (BFGS)算法。尽管对于优化来说,这些近似算法运作良好,但是将估计参数的估计协方差矩阵基于这些近似方法应该小心谨慎(如 Thisted, 1987)。

哈伯曼(Haberman, 1989)将牛顿-拉夫逊算法用于潜类模型,潘和汤普逊(Pan and Thompson, 2003)将其用于广义线性混和模型,拉贝-赫斯基思等人(Rabe-Hesketh et al., 2002, 2004a)将其用于广义线性潜混和模型。潘和汤普逊使用对数似然的解析一阶和二阶导数,而拉贝-赫斯基思等人则采用数值导数。麦克拉奇(McCulloch, 1997)以及库克和程(Kuk 和 Cheng, 1997)针对广义线性混和模型讨论了蒙特卡洛牛顿-拉夫逊算法。

费舍尔计分算法被朗福特(Longford, 1987)用于线性混和模型。$\mathrm{BH}^3$ 算法被 Arminger and Küsters (1989)、Skrondal (1996)以及 Hedeker and Gibbons (1994, 1996a)和其他研究者用于潜变量模型。杰拉斯科格 (Jöreskog, 1967)将 Davidon-Fletcher-Powell (DFP)算法引入因子模型。司考戈(Skaug, 2002)用准牛顿方法处理广义线性模型的线搜索。其中,一阶导数由自动差分获得,即计算导数的代码由计算程序从计算对数似然的程序中自动生成。

## 6.5 非参数最大似然估计

既然似然是**有限混和**的,并无积分包含在内,因此,对包含离散潜变量的模型进行估计的最直接的方法是 EM 或梯度算法。离散分布用位置有限集 $e_c, c=1,\cdots,C$ 和概率或质量 $\pi_c$(在这些位置处)来表示。如果选择离散分布的质量数 $C$ 来最大化似然,则可得非参最大似然估计量(如 Simar,1976;Laird,1978;Lindsay,1983),见 4.4.2 节。尝试增加额外的质量点会导致一个趋于零的估计概率或两个几乎一致的估计位置。

在本节我们主要介绍寻找非参最大似然估计量(NPMLE)的质量数的方法。一个常用的方法是在似然最大化的过程中(如 Butler and Louis,1992),使用大量的质量点做为起始估计,并忽略融合于其他点或者质量接近 0 的点。另一个方法是使用方向导数的概念(如 Simar,1976;Jewell,1982;Böhning,1982;Lindsay,1983;Rabe-Hesketh et al.,2003a),一个接一个地引入质量点,即赫克曼和辛格尔(Heckman and Singer,1984)所称的加托导数(Gateaux derivative)。

考虑对 $C$ 个质量具有最大化对数似然 $l(\hat{\boldsymbol{\vartheta}}^C,\hat{\boldsymbol{\pi}}^C,\hat{\boldsymbol{e}}^C)$ 的单个潜变量模型。为确定这是否为 NPMLE 解,我们考虑沿位置为 $(\hat{\mathbf{e}}^C,\boldsymbol{e}^{C+1})'$ 的路径 $([1-\lambda]\hat{\boldsymbol{\pi}}^C,\lambda)'$,对离散质量点分布进行变换,其中 $\lambda=0$ 对应于当前解,$\lambda=1$ 将单元质量处于新位置 $e^{C+1}$,则方向导数可被定义为

$$\Delta(e^{C+1}) = \lim_{\lambda\to 0}\frac{l(\hat{\boldsymbol{\vartheta}}^C,([1-\lambda]\hat{\boldsymbol{\pi}}^C,\lambda)',(\hat{\mathbf{e}}^C,e^{C+1})') - l(\hat{\boldsymbol{\vartheta}}^C,\hat{\boldsymbol{\pi}}^C,\hat{\mathbf{e}}^C)}{\lambda} \quad (6.16)$$

根据广义混和最大似然理论(Lindsay,1983;Böhning,1982),当且仅当对于所有 $e^{C+1}$,$\Delta(e^{C+1})\leqslant 0$ 时,可找到 NPMLE。

拉贝-赫斯基思等(Rabe-Hesketh,2003a)提出在一个能够生成广泛值域的优质网格范围内搜寻,并且如果对于一个小量 $\lambda$,式(6.16)的分子对于所有位置都为负时,算法结束。这种方法类似于由希玛(Simar,1976)提出,经赫克曼和辛格尔(Heckman and Singer,1984),弗曼和兰伯特(Follmann and Lambert,1989)以及其他人所改进的算法。林赛(Lindsay,1995)和波宁(Böhning,2000)介绍了寻找 NPMLE(包括质量数和参数估计)的算法。

NPMLE 的重要优点在于我们无须为潜变量假定一个参数分布,这潜在地使推论更加稳健。尽管如此,对于包含协变量测量误差的广义线性模型来说,模拟研究表明,基于多元正态潜变量的推论很容易陷入设定错误(如 Thoresen and Laake,2000)。尽管沙弗尔(Schafer,2001)和拉贝-赫斯基思等(Rabe-Hesketh et al.,2003a)发现,在假定常规(设定错误)模型有偏的情况下进行估计,这些估计的误差均方根会小于无偏的 NPML 估计。对于具有大量潜变量的模型的 NPMLE 的执行效率知之甚少。对于类别响应模型,在位置趋近于 $\pm\infty$ 时,边界解将会引起一些问题。

非参最大似然估计被用于生存或存续模型(如 Heckman and Singer,1984;Holmas,2002),题器响应模型(如 de Leeuw and Verhelst,1986;Lindsay et al.,1991),包含协变量测量误差的广义线性模型(如 Roeder et al.,1996;Aitkin and Rocci,2002;Rabe-Hesketh et al.,2003a),随机系数模型(如 Davies and Pickles,1987;Aitkin,1999a)以及元分析(如 Aitkin,1999b)。在本书的许多应用研究中都使用了 NPMLE,如 9.5 节、11.2 节、11.3.3 节和 14.2 节。

## 6.6 受约束/残差最大似然(REML)

首先,我们有必要考虑如下的简单线性回归模型

$$y_i = \mathbf{x}_i'\boldsymbol{\beta} + \epsilon_i, \quad \epsilon_i \sim \mathrm{N}(0,\theta), \quad i = 1,\cdots,N$$

其中,$\boldsymbol{\beta}$ 包含 $P$ 个回归参数。残方差的最大似然估计量为

$$\hat{\theta} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}})^2$$

众所周知,该估计量向下偏移一个"偏误因子"$\frac{N-P}{N}$,即 $\mathrm{E}(\hat{\theta}) = \frac{N-P}{N}\theta$。当回归参数 $\boldsymbol{\beta}$ 已知时,$\hat{\theta}$ 无偏,但是由于 $\mathbf{x}_i'\hat{\boldsymbol{\beta}}$ 比 $\mathbf{x}_i'\boldsymbol{\beta}$"更加拟合数据",故基于 $\boldsymbol{\beta}$ 的 $\hat{\theta}$ 则会有偏。因此,通常采用偏误校正估计量

$$\hat{\theta} = \frac{1}{N-P}\sum_{i=1}^{N}(y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}})^2$$

对于潜变量模型来说,方差参数的估计也预期向下偏移,因此存在同样的有偏问题。如,对于一个双层次随机截距模型,劳登布什和布莱克(Raudenbush and Bryk,2002)指出,随机截距方差 $\psi$ 的最大似然估计量有偏且其近似偏误因子为$\frac{J-P}{J}$。

为处理这一问题,帕特森和汤普逊(Patterson and Thompson,1971)提出了所谓的受约束或残差最大似然方法(REML)。在这种情况下,最大似然并不直接应用于响应 $\mathbf{y}$,而是响应的线性函数或"**误差对比度**(error contrasts)"$\mathbf{A}_\mathbf{y}$。重要的是,$\mathbf{A}$ 被设定为正交于 $\mathbf{X}$,因此 $\mathbf{A}_\mathbf{y}$ 会清除来源于模型的固定效应。REML 本身并不产生固定效应 $\boldsymbol{\beta}$ 的估计。

REML 也可由贝叶斯方法推导出来(见 6.11 节)。具体来说,就是对 $\boldsymbol{\beta}$ 设定一个扁平先验,而变量和协方差参数则被视为固定的。在积分出 $\boldsymbol{\beta}$ 和潜变量后,后者参数可用最大边缘似然来估计。经验贝叶斯(见 7.3.1 节)被用于计算潜变量"得分"和估计 $\boldsymbol{\beta}$ 参数(如 Harville,1977;Dempster et al.,1981;Laird and Ware,1982)。

尽管为只包含连续响应的线性混和模型而发展出来,基于罚伪似然的近似 REML 方法也被用于广义线性混和模型(如 Schall,1991;Breslow and Clayton,1993;McGilchrist,1994;Stiratelli et al.,1984)。朗福特(Longford,1993:236)提出了另一种方法,给边缘对数似然加上一个罚项。

当对 REML 和最大似然(ML)的结果进行对比时,很难看出谁是优胜者。有关 REML 胜过 ML 的普遍的观点是 REML 可产生方差和协方差参数的无偏估计量。然而,值得注意的是,只有在相对于固定效应数量,群组的数量较少时,ML 的偏误才显得突出。在这种情况下,潜变量模型本身的效果都是值得怀疑的,因此,ML 与 REML 的结果优劣变得次要。此外,均方差通常取代偏误被用作最佳性准则。有意思的是,均方差可能大于 REML(如 Corbeil and Searle,1976),巴辛(Busing,1993)以及范德礼顿和巴辛(van der Leeden and Busing,1994)的模拟结果也表明了这一点。REML 的一个缺点是,由于 REML 本身不提供固定效应的估计,因而对固定参数的偏差检验被排除在外。另一方面,对于平衡混合 ANOVA 模型这种特殊例子,对方差和协方差的 REML 估计等于经典 ANOVA 矩估计量。在这种情况下,这将导致 REML 估计量具有最小方差属性并且不依赖于任何正态假设。最后,对于异常值处理,REML 被认为不如 ML 灵敏(Verbyla,1993)。

## 6.7 有限信息方法

本节中我们考虑一种具有固定群组大小，$n_j = n, j = 1, \cdots, J$ 的条件多元正态潜响应 $\mathbf{y}_j^*$ 的模型。尽管当顶层单元中的第一层级单元为常数时，更高层次的模型更适用，但在这里我们只考虑双层次模型作为**多元增长曲线模型**。此外，我们假定潜变量 $\boldsymbol{\zeta}$ 是多元正态的，因此，潜响应的边缘分布（关于潜变量）也是多元正态的。在这种情况下，简化式参数是表征边缘均值和协方差结构的参数，见 4.9.2 节。与多元正态一样，一元分布和二元分布（关于其他潜响应的边缘）也是正态的。

在有限信息方法中，我们首先用一元分布和二元分布来估计简化式参数的"经验的"或无约束形式。例如，对于一个未包含协变量的模型，这是基本不满足模型隐含的约束条件的 $\mathbf{y}_j^*$ 的无约束均值和协方差（在二项情况下是四项相关）。然后我们使用式(6.1)的加权最小二乘拟合函数来估计结构参数，使模型隐含的与所估计的（无约束）简化式参数之间的距离最小化。对于 $\mathbf{y}_j = \mathbf{y}_j^*$ 处的连续响应，一元分布和二元分布包含有关简化式参数的所有信息。而在粗略响应中（如二项响应），这种情况不属于所谓的"有限信息"。

在本文中，使用一元和二元信息的理念源自克里斯托弗森（Christofferson，1975），并被穆森在一系列论文中进行了扩展（如 Muthén，1978，1981，1982，1983，1984，1988a，1989bc）。由于有限信息方法在计量心理学外鲜有人知，我们在此提供一定程度的概览。

考虑穆森（Muthén，1983，1984）介绍的模型，该模型扩展了第 3 章所提出的潜响应 $\mathbf{y}^*$ 的结构方程模型。方程(3.33)给出的结构模型为

$$\boldsymbol{\eta}_j = \boldsymbol{\alpha} + \mathbf{B}\boldsymbol{\eta}_j + \boldsymbol{\Gamma}\mathbf{x}_j + \boldsymbol{\zeta}_j$$

响应模型为

$$\mathbf{y}_j^* = \boldsymbol{\nu} + \boldsymbol{\Lambda}\boldsymbol{\eta}_j + \mathbf{K}\mathbf{x}_j + \boldsymbol{\epsilon}_j$$

对于连续响应，观测响应则等于潜响应。二项响应、定序响应和删失观测响应通过 2.4 节所述的阈值函数与潜响应相关。简化式则变为

$$\mathbf{y}_j^* = \boldsymbol{\nu} + \boldsymbol{\Lambda}(\mathbf{I} - \mathbf{B})^{-1}[\boldsymbol{\alpha} + \boldsymbol{\Gamma}\mathbf{x}_j + \boldsymbol{\zeta}_j] + \mathbf{K}\mathbf{x}_j + \boldsymbol{\epsilon}_j$$

伴随期望结构

$$\mathrm{E}(\mathbf{y}_j^* \mid \mathbf{x}_j) = \underbrace{\boldsymbol{\nu} + \boldsymbol{\Lambda}(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\alpha}}_{\boldsymbol{\Pi}_0} + \underbrace{[\boldsymbol{\Lambda}(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Gamma} + \mathbf{K}]}_{\boldsymbol{\Pi}_1}\mathrm{x}_j$$

和协方差结构

$$\boldsymbol{\Omega} \equiv \mathrm{Cov}(\mathbf{y}_j^* \mid \mathbf{x}_j) = \boldsymbol{\Lambda}(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Psi}(\mathbf{I} - \mathbf{B})^{-1\prime}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}$$

为简单起见，我们从现在开始假定观测响应是二分的。在这种情况下，为了辨识将 $\boldsymbol{\Theta}$ 或 $\boldsymbol{\Omega}^*$ 的对角元素固定为 1。沿用穆森的方法，我们在本节中用二级参数化，因此，协方差结构变为

$$\boldsymbol{\Omega}^* = \mathrm{diag}(\boldsymbol{\Omega})^{-\frac{1}{2}}\boldsymbol{\Omega}\mathrm{diag}(\boldsymbol{\Omega})^{-\frac{1}{2}}$$

此外，为了辨识可假定阈值为零。为简化之后的推衍，我们定义增广协变量向量 $\mathbf{z}_j = (1, \mathbf{x}_j')'$，除了协变量外还包括截距 1。期望结构则可写为

$$\mathrm{E}(\mathbf{y}_j^* \mid \mathbf{z}_j) = \boldsymbol{\Pi}\mathbf{z}_j$$

其中，$\boldsymbol{\Pi} = (\boldsymbol{\Pi}_0, \boldsymbol{\Pi}_1)$ 为包括截距的简化式回归矩阵。沿用穆森的方法，我们设定 $\boldsymbol{\zeta}_j$ 和 $\boldsymbol{\epsilon}_j$ 服从多项式分布，得

$$\mathbf{y}_j^* \mid \mathbf{z}_j \sim \mathrm{N}_n(\boldsymbol{\Pi}\mathbf{z}_j, \boldsymbol{\Omega}^*)$$

我们可将变量或响应 $i, y_{ij}^*, i=1,2,\cdots,n$,的一元分布写为

$$y_{ij}^* \mid \mathbf{z}_j \sim \mathrm{N}(\boldsymbol{\pi}_i \mathbf{z}_j, 1)$$

其中,$\boldsymbol{\pi}_i$ 为 $\boldsymbol{\Pi}$ 的第 $i$ 行,则两个响应 $y_{ij}^*$ 和 $y_{i'j}^*, i, i'=1,2,\cdots,n, i \neq i'$的二元分布可写为

$$y_{ij}^*, y_{i'j}^* \mid \mathbf{z}_j \sim \mathrm{N}_2\left(\begin{bmatrix} \boldsymbol{\pi}_i \mathbf{z}_j \\ \boldsymbol{\pi}_{i'} \mathbf{z}_j \end{bmatrix}, \begin{bmatrix} 1 & \\ \omega_{ii'}^* & 1 \end{bmatrix}\right)$$

其中,$\omega_{ii'}^*$ 为 $\boldsymbol{\Omega}^*$ 的第 $ii'$ 个残相关单元。

现在我们可清楚地说明穆森提出的三阶段有限信息方法。注意 $\boldsymbol{\Pi}$ 和 $\boldsymbol{\Omega}^*$ 代表简化式参数的无约束格式。

**第一阶段**:第一个估计阶段由一元信息得出简化式截距和回归参数 $\boldsymbol{\Pi}$ 的有限信息最大似然估计。对每一个 $i$,设定其一元正态概率回归模型为

$$\Pr(y_{ij}=1 \mid \mathbf{z}_{ij}) = \Phi_1(\boldsymbol{\pi}_i \mathbf{z}_j)$$

其中,$\Phi_1(\cdot)$为标准正态累积分布函数。群组 j 和变量 i 的对数似然贡献就成为

$$l_i^j(\boldsymbol{\pi}_i) \equiv y_{ij} \ln \Phi_1(\boldsymbol{\pi}_i \mathbf{z}_j) + (1-y_{ij}) \ln[1-\Phi_1(\boldsymbol{\pi}_i \mathbf{z}_j)]$$

对于每一个 $i$,一元对数似然 $\sum_{j=1}^{J} l_i^j(\boldsymbol{\pi}_i)$ 即被最大化,得出一致估计 $\hat{\boldsymbol{\pi}}_i$。群组 $j$ 的梯度为

$$\mathbf{g}_1^j = \left[\frac{\partial l_1^j}{\partial \boldsymbol{\pi}_1}, \frac{\partial l_2^j}{\partial \boldsymbol{\pi}_2}, \cdots, \frac{\partial l_n^j}{\partial \boldsymbol{\pi}_n}\right]'$$

**第二阶段** a:依赖于第一阶段所得到的简化式截距和回归参数估计 $\hat{\boldsymbol{\Pi}}$,$\boldsymbol{\Omega}^*$ 中的简化式残相关用有限信息“伪”最大似然来估计,该估计基于每一对响应 $ii'i>i'$的二元信息,即

$$\Pr(y_{ij}=1, y_{i'j}=1 \mid \mathbf{z}_{ij}, \mathbf{z}_{i'j}) = \Phi_2(\boldsymbol{\pi}_i \mathbf{z}_j, \boldsymbol{\pi}_{i'} \mathbf{z}_j, \omega_{ii'}^*)$$
$$\Pr(y_{ij}=0, y_{i'j}=1 \mid \mathbf{z}_{ij}, \mathbf{z}_{i'j}) = \Phi_2(-\boldsymbol{\pi}_i \mathbf{z}_j, \boldsymbol{\pi}_{i'} \mathbf{z}_j, -\omega_{ii'}^*)$$
$$\Pr(y_{ij}=1, y_{i'j}=0 \mid \mathbf{z}_{ij}, \mathbf{z}_{i'j}) = \Phi_2(\boldsymbol{\pi}_i \mathbf{z}_j, -\boldsymbol{\pi}_{i'} \mathbf{z}_j, -\omega_{ii'}^*)$$
$$\Pr(y_{ij}=0, y_{i'j}=0 \mid \mathbf{z}_{ij}, \mathbf{z}_{i'j}) = \Phi_2(-\boldsymbol{\pi}_i \mathbf{z}_j, -\boldsymbol{\pi}_{i'} \mathbf{z}_j, \omega_{ii'}^*)$$

其中,$\Phi_2(\mu_1, \mu_2, \rho)$为均值 $\mu_1$ 和 $\mu_2$、相关度 $\rho$ 以及残相关 $\omega_{ii'}^*$ 的二元标准正态累积分布函数。给定由第一阶段所得出的估计 $\hat{\boldsymbol{\pi}}_i$ 和 $\hat{\boldsymbol{\pi}}_{i'}$,对应的二元对数似然贡献变为

$$\begin{aligned} l_{ii'}^j(\omega_{ii'}^* \mid \hat{\boldsymbol{\pi}}_i \hat{\boldsymbol{\pi}}_{i'}) = {} & y_{ij} y_{i'j} \ln \Phi_2(\hat{\boldsymbol{\pi}}_i \mathbf{z}_j, \hat{\boldsymbol{\pi}}_{i'} \mathbf{z}_j, \omega_{ii'}^*) + \\ & y_{ij}(1-y_{i'j}) \ln \Phi_2(\hat{\boldsymbol{\pi}}_i \mathbf{z}_j, -\hat{\boldsymbol{\pi}}_{i'} \mathbf{z}_j, -\omega_{ii'}^*) + \\ & (1-y_{ij}) y_{i'j} \ln \Phi_2(-\hat{\boldsymbol{\pi}}_i \mathbf{z}_j, \hat{\boldsymbol{\pi}}_{i'} \mathbf{z}_j, -\omega_{ii'}^*) + \\ & (1-y_{ij})(1-y_{i'j}) \ln \Phi_2(-\hat{\boldsymbol{\pi}}_i \mathbf{z}_j, -\hat{\boldsymbol{\pi}}_{i'} \mathbf{z}_j, \omega_{ii'}^*) \end{aligned}$$

对于每一对 $ii'$,“伪对数似然”(根据 Parke,1986 的含义)$\sum_{i=1}^{J} l_{ii'}^j(\omega_{ii'}^* \mid \hat{\boldsymbol{\pi}}_i, \hat{\boldsymbol{\pi}}_{i'})$ 则被最大化,得到一致估计 $\hat{\omega}_{ii'}^*$。由最大伪似然所得出的群组 j 的梯度为

$$\mathbf{g}_2^j = \left[\frac{\partial l_{21}^j}{\partial \omega_{21}^*}, \frac{\partial l_{31}^j}{\partial \omega_{31}^*}, \frac{\partial l_{32}^j}{\partial \omega_{32}^*}, \cdots, \frac{\partial l_{n,n-1}^j}{\partial \omega_{n,n-1}^*}\right]'$$

为方便以后使用,我们定义

$$\mathbf{g}^j = \left[\mathbf{g}_1^{j\prime}, \mathbf{g}_2^{j\prime}\right]'$$

$$\mathbf{g} = \sum_{j=1}^{N} \mathbf{g}^j$$

并将最大伪似然计算出的关于简化式截距和回归参数的二元对数似然的梯度放到向量中,即

$$\mathbf{g}_{21}^{j} = \left[\frac{\partial l_{21}^{j}}{\partial \boldsymbol{\pi}_1}, \frac{\partial l_{21}^{j}}{\partial \boldsymbol{\pi}_2}, \frac{\partial l_{31}^{j}}{\partial \boldsymbol{\pi}_1}, \frac{\partial l_{31}^{j}}{\partial \boldsymbol{\pi}_3}, \cdots, \frac{\partial l_{n,n-1}^{j}}{\partial \boldsymbol{\pi}_{n-1}}, \frac{\partial l_{n,n-1}^{j}}{\partial \boldsymbol{\pi}_n}\right]'$$

中。

**第二阶段** b:将简化式参数 $\mathbf{\Pi}$ 和 $\mathbf{\Omega}^*$ 的非冗余元素构成向量 $\boldsymbol{\sigma}$,则估计的简化式参数的估计渐近协方差矩阵,$\widehat{\text{Cov}}(\hat{\boldsymbol{\sigma}})$,可由第一阶段和第二阶段 $a$ 的边缘信息导出(如 Lee,1982;Muthén,1984)。

使用中值定理,将最大似然估计的梯度 $\mathbf{g}(\hat{\boldsymbol{\sigma}})$ 围绕真实值 $\bar{\boldsymbol{\sigma}}$ 展开,得

$$\mathbf{0} = \mathbf{g}(\hat{\boldsymbol{\sigma}}) = \mathbf{g}(\bar{\boldsymbol{\sigma}}) + \frac{\partial \mathbf{g}(\boldsymbol{\sigma}^*)}{\partial \boldsymbol{\sigma}}(\hat{\boldsymbol{\sigma}} - \bar{\boldsymbol{\sigma}})$$

其中,$\boldsymbol{\sigma}^*$ 是 $\hat{\boldsymbol{\sigma}}$ 和 $\bar{\boldsymbol{\sigma}}$ 中间的一些点,乘以 $J^{\frac{1}{2}}$,可得

$$J^{\frac{1}{2}}(\hat{\boldsymbol{\sigma}} - \bar{\boldsymbol{\sigma}}) = \left(-J^{-1}\frac{\partial \mathbf{g}(\boldsymbol{\sigma}^*)}{\partial \boldsymbol{\sigma}}\right)^{-1} \times J^{-\frac{1}{2}}\mathbf{g}(\bar{\boldsymbol{\sigma}})$$

对于第一项,根据大数定理可得

$$J^{-1}\frac{\partial \mathbf{g}(\boldsymbol{\sigma}^*)}{\partial \boldsymbol{\sigma}} = J^{-1}\sum_{j=1}^{J}\frac{\partial \mathbf{g}^j(\boldsymbol{\sigma}^*)}{\partial \boldsymbol{\sigma}} \xrightarrow{\text{P}} \lim_{J\to\infty} J^{-1}\sum_{j=1}^{J}\text{E}\left(\frac{\partial \mathbf{g}_j(\bar{\boldsymbol{\sigma}})}{\partial \boldsymbol{\sigma}}\right) \equiv \mathbf{A}$$

其中,$\xrightarrow{\text{P}}$代表依概率收敛。注意,由于不含任何相关,故关于相关度的一元似然的梯度的导数为0,因此,$\mathbf{A}$ 可分块为

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}$$

其中,$\mathbf{A}_{12} = \mathbf{0}$。

对于第二项,根据多元中心极限定理可得

$$J^{-\frac{1}{2}}\mathbf{g}(\bar{\boldsymbol{\sigma}}) = J^{-\frac{1}{2}}\sum_{j=1}^{J}\mathbf{g}^j(\bar{\boldsymbol{\sigma}}) \xrightarrow{\text{D}} \text{N}(\mathbf{0}, \mathbf{V})$$

其中,$\xrightarrow{\text{D}}$表示依分布收敛,并且由于 $\sum_{j=1}^{J}\text{E}(\mathbf{g}^j(\bar{\boldsymbol{\sigma}})) = \mathbf{0}$,则预期得分在最大处为零,因此

$$\mathbf{V} \equiv \lim_{J\to\infty}\text{Cov}(J^{-\frac{1}{2}}\mathbf{g}(\bar{\boldsymbol{\sigma}})) = \lim_{J\to\infty}J^{-1}\sum_{j=1}^{J}\text{E}(\mathbf{g}^j(\bar{\boldsymbol{\sigma}})\mathbf{g}^j(\bar{\boldsymbol{\sigma}})')$$

从而可得

$$J^{\frac{1}{2}}(\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}) \xrightarrow{\text{D}} \text{N}(\mathbf{0}, \mathbf{A}^{-1}\mathbf{V}\mathbf{A}^{-1\prime})$$

在正确的模型设定下,根据"信息矩阵等价性"可得

$$\lim_{J\to\infty}J^{-1}\sum_{j=1}^{J}\text{E}\left(\frac{\partial \mathbf{g}^j(\boldsymbol{\sigma})}{\partial \boldsymbol{\sigma}}\right) = -\lim_{J\to\infty}J^{-1}\sum_{j=1}^{J}\text{E}(\mathbf{g}^j(\boldsymbol{\sigma})\mathbf{g}^j(\boldsymbol{\sigma})')$$

根据上述结果,我们可以用前几个阶段所得到的梯度来估计 A 中的子矩阵为

$$\hat{\mathbf{A}}_{11} = -J^{-1}\sum_{j=1}^{J}\mathbf{g}_1^j\mathbf{g}_1^{j\prime}$$

$$\hat{\mathbf{A}}_{22} = -J^{-1}\sum_{j=1}^{J}\mathbf{g}_2^j\mathbf{g}_2^{j\prime}$$

且

$$\hat{\mathbf{A}}_{21} = -J^{-1}\sum_{j=1}^{J}\mathbf{g}_{21}^{j}\mathbf{g}_{1}^{j'}$$

根据梯度的经验协方差可估计协方差矩阵 V，即

$$\hat{\mathbf{V}} = J^{-1}\sum_{j=1}^{J}\mathbf{g}^{j}\mathbf{g}^{j'}$$

所估计的简化式参数的渐近协方差矩阵最终可估计为

$$\hat{\mathbf{W}} \equiv \widehat{\mathrm{Cov}}(\hat{\boldsymbol{\sigma}}) = \hat{\mathbf{A}}^{-1}\hat{\mathbf{V}}\hat{\mathbf{A}}^{-1\prime}$$

**第三阶段**：简化式参数被视为基本参数 $\boldsymbol{\vartheta}$ 的函数 $\boldsymbol{\sigma}(\boldsymbol{\vartheta})$。一致估计量 $\hat{\boldsymbol{\vartheta}}$ 可由 $\boldsymbol{\sigma}(\boldsymbol{\vartheta})$ 对 $\hat{\boldsymbol{\sigma}}$ 的拟合得到，并使加权最小二乘(WLS)标准最小化

$$\mathrm{F}(\boldsymbol{\vartheta}) = \frac{1}{2}[\boldsymbol{\sigma}(\boldsymbol{\vartheta}) - \hat{\boldsymbol{\sigma}}]'\hat{\mathbf{W}}^{-1}[\boldsymbol{\sigma}(\boldsymbol{\vartheta}) - \hat{\boldsymbol{\sigma}}] \tag{6.17}$$

定义

$$\boldsymbol{\Delta} \equiv \frac{\partial\boldsymbol{\sigma}(\hat{\boldsymbol{\vartheta}})}{\partial\boldsymbol{\vartheta}}$$

则 $\hat{\boldsymbol{\vartheta}}$ 的渐近协方差矩阵的基于模型的估计量为

$$\widehat{\mathrm{Cov}}(\hat{\boldsymbol{\vartheta}}) = J^{-1}(\boldsymbol{\Delta}'\hat{\mathbf{W}}^{-1}\boldsymbol{\Delta})^{-1}$$

当模型设定正确时，这个估计量是一致的。可得到针对所估计的简化式参数的绝对拟合大样本卡方分布检验统计量 $2JF(\hat{\boldsymbol{\vartheta}})$。

关于技术细节，请参考库斯特(Küsters，1987)以及 Muthén and Satorra(1996)。对于前两个阶段的多项相关，奥尔森(Olsson，1979)提供了更多细节，奥尔森等(Olsson et al.，1982)针对多序列相关，穆森(Muthén，1989c)针对托比(Tobit)相关进行了分析。表 4.2 给出了库斯特(Küsters，1987)对各种潜响应相关所做的综述，他将其标记为"多二列(polytobiserial)"。

重要的是，穆森(Muthén，1993)提出了一种类似的稳健正态理论估计方法(见萨托拉的综述(Satorra，1990))，并用于二项响应以及广义穆森模型(Muthén et al.，1997)。$\hat{\boldsymbol{\vartheta}}$ 的"稳健"渐近协方差矩阵为

$$\widehat{\mathrm{Cov}}(\hat{\boldsymbol{\vartheta}}) = J^{-1}(\boldsymbol{\Delta}'\widehat{\mathbf{W}}^{-1}\boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}'\widehat{\mathbf{W}}^{-1}\hat{\mathbf{W}}\widehat{\mathbf{W}}^{-1}\boldsymbol{\Delta}(\boldsymbol{\Delta}'\widehat{\mathbf{W}}^{-1}\boldsymbol{\Delta})^{-1}$$

穆森(Muthén，1993)提出在上述表达式中可简单地用 $\widehat{\mathbf{W}} = \mathbf{I}$，就像在 *WLS* 标准式(6.17)中使用的那样，可有效地简化后者的未加权最小二乘。穆森等(Muthén et al.，1997)将 $\widehat{\mathbf{W}}$ 设定为一个以估计方差 $\hat{\boldsymbol{\sigma}}$ 为元素的对角矩阵。所得到的 $\widehat{\mathbf{W}}$ 被用于'稳健'协方差矩阵，从而拟合标准成为对角线上的加权最小二乘。这些方法的一个好处是 $\widehat{\mathbf{W}}$ 无须反转，而这对于"大"模型/或小样本/或高度偏斜的二项响应会存在一些问题。萨托拉和本特勒(如 Satorra and Bentler，1994；Satorra，1992)更进一步提出了一种绝对拟合的"稳健"检验。另外，穆森等(Muthén et al.，1997)讨论了这些方法与 6.9 节中所介绍的 GEE 方法之间的联系。

穆森和其他人所发展的有限信息方法具有很多优点。它可处理一般模型框架，尽管仅指包含多元正态潜响应的模型。因而，对于那些包含对数概率链接或泊松链接的模型则不适用。这个方法在计算上非常有效，将高维积分问题简化为一系列一元或二元积

分，这对于具有多个潜变量的模型来说非常有价值。这个方法看起来在估计上也非常有效，其估计结果非常接近于最大似然的估计。其主要的局限性则在于，只能通过多组模型来处理缺失数据问题，因而在实践中，仅有少数几种缺失数据模式可以得到处理。对于样本规模较小的复杂模型，蒙特卡洛实验（如 Muthén and Kaplan，1992）表明这个方法效果很差。

## 6.8 最大伪似然估计

我们首先讨论针对广义线性模型的迭代重加权最小二乘（IRLS）算法以及针对多元线性模型的迭代广义最小二乘（IGLS）算法。这不仅因为它们本身有意义，并且也因为它们是伪似然估计、边缘和罚伪似然估计（MQL 和 PQL）以及广义估计方程（GEE）的前驱。

### 6.8.1 迭代重加权最小二乘

考虑具有下述对数似然的广义线性模型（见 2.2 节）

$$l = \sum_i [y_i\theta_i - b(\theta_i)]/\phi + c(y_i,\phi)$$

在这个情形中，对于 $p=1,\cdots,P$，似然方程为

$$\frac{\partial l}{\partial \beta_p} = \sum_i \left[y_i \frac{\partial\theta_i}{\partial\mu_i}\frac{\partial\mu_i}{\partial\beta_p} - \frac{\partial b(\theta_i)}{\partial\theta_i}\frac{\partial\theta_i}{\partial\mu_i}\frac{\partial\mu_i}{\partial\beta_p}\right]/\phi = 0$$

对于累积量函数 $b(\cdot)$，则

$$\frac{\partial b(\theta_i)}{\partial\theta_i} = \mu_i$$

$$\frac{\partial^2 b(\theta_i)}{\partial\theta_i^2} = \frac{\partial\mu_i}{\partial\theta_i} = \left[\frac{\partial\theta_i}{\partial\mu_i}\right]^{-1} = V(\mu_i)$$

将这些表达式代入似然方程

$$\frac{\partial l}{\partial\beta_p} = \sum_i \frac{\partial\mu_i}{\partial\beta_p}[\phi V(\mu_i)]^{-1}[y_i - \mu_i] = 0 \tag{6.18}$$

引入关系 $\mu_i = g^{-1}(v_i)$ 和 $\nu_i = \mathbf{x}_i'\boldsymbol{\beta}$，该方程可近一步简化为

$$\frac{\partial\mu_i}{\partial\beta_p} = \frac{\partial\mu_i}{\partial\nu_i}x_{pi} = \frac{\partial g^{-1}(\nu_i)}{\partial\nu_i}x_{pi} = \frac{x_{pi}}{g'(\mu_i)}$$

其中，$g'(\mu_i)$ 为 $g(\cdot)$ 在 $\mu_i$ 处的一阶导数。

将这些表达式代入似然方程

$$\frac{\partial l}{\partial\beta_p} = \sum_i \frac{x_{pi}}{g'(\mu_i)\phi V(\mu_i)}[y_i - \mu_i] = 0 \tag{6.19}$$

对于可能具有非齐次残方差 $\phi_i = \sigma_i^2$ 的线性模型（同一性链接且正态分布），$V(\mu_i)=1$，$g'(\mu_i)=1$，使得似然方程在 $\boldsymbol{\beta}$ 上为线性，即

$$\frac{\partial l}{\partial\beta_p} = \sum_i \frac{x_{pi}}{\sigma_i^2}[y_i - \mathbf{x}_i'\boldsymbol{\beta}] = 0 \tag{6.20}$$

并且可用权重为 $1/\sigma_i^2$ 的**加权最小二乘**求解

$$\hat{\beta} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \tag{6.21}$$

其中，$\mathbf{V}$ 是对角元素等于 $\sigma_i^2$ 的对角矩阵。

迭代重加权最小二乘（IRLS）是在每个迭代步对似然方程进行线性化的过程，因此，

下一步迭代的估计就可由加权最小二乘得出。设 $\beta^k$ 表示“当前”迭代的估计，$\mu_i^k$ 为对应的均值。**工作变量**(working variates) $z_i^k$ 定义为

$$z_i^k = g(\mu_i^k) + [y_i - \mu_i^k]g'(\mu_i^k)$$

使得

$$y_i - \mu_i^k = [z_i^k - \mathbf{x}_i'\boldsymbol{\beta}^k]/g'(\mu_i^k)$$

现在可知，可以用权值为 $1/\sigma_i^2 = [g'(\mu_i^k)^2\phi V(\mu_i^k)]^{-1}$ 的加权最小二乘来(假定模型为线性)校正估计。将这些权值带入加权最小二乘方程(6.20)

$$0 = \sum_i \frac{x_{pi}}{g'(\mu_i^k)^2\phi V(\mu_i^k)}[z_i - \mathbf{x}_i'\boldsymbol{\beta}] = \sum_i \frac{x_{pi}}{g'(\mu_i^k)\phi V(\mu_i^k)} \frac{[z_i - \mathbf{x}_i'\boldsymbol{\beta}]}{g'(\mu_i^k)}$$

$$= \sum_i \frac{x_{pi}}{g'(\mu_i^k)\phi V(\mu_i^k)}[y_i - \mu_i]$$

我们得到一开始的似然方程(6.19)，除了第一项(权重)的分母被固定为由上一次迭代 $k$ 所得到的估计。用加权最小二乘求解这个方程得到估计 $\boldsymbol{\beta}^{k+1}$，即得到新的权重，那么新的估计就用这个“重加权的”最小二乘来继续求解，如此往复，直至迭代收敛。注意对于广义线性模型，迭代重加权最小二乘算法等于费舍尔计分。

对该算法的另一个概念化方法是用一个线性模型来近似广义线性模型，而该线性模型使用一阶泰勒级数展开。令 $h(\nu_i) \equiv g^{-1}(\nu_i) = \mu_i$，以及 $h'(\nu_i)$ 为 $\nu_i$ 处的一阶导数。在第 $k$ 次迭代中，$y_i$ 近似为

$$y_i = h(\nu_i^k) + \mathbf{x}_i'(\boldsymbol{\beta} - \boldsymbol{\beta}^k)h'(\nu_i^k) + \epsilon$$

其中，$\text{Var}(\epsilon) = \phi V(\mu_i)$。移项得

$$y_i - h(\nu_i^k) + \mathbf{x}_i'\boldsymbol{\beta}^k h'(\nu_i^k) = \mathbf{x}_i'\boldsymbol{\beta}h'(\nu_i^k) + \epsilon$$

乘以 $1/h'(\nu_i^k) = g'(\mu_i^k)$，可得

$$[y_i - \mu_i^k]g'(\mu_i^k) + \mathbf{x}_i'\boldsymbol{\beta}^k = z_i^k = \mathbf{x}_i'\boldsymbol{\beta} + g'(\mu_i^k)\epsilon$$

即是均值为 $\mathbf{x}_i'\boldsymbol{\beta}$，方差为 $g'(\mu^k)^2\phi V(\mu)$ 的 $z_i^k$ 的线性模型。

注意，该算法的细节仅仅取决于广义线性模型的链接和方差函数。如果我们希望设定一个任意的链接和方差函数，可以采用同样的算法估计参数，即便设定与任何统计模型都毫不对应。这种无须设定模型的参数估计就被称为**伪似然**(quasi-likelihood)(Wedderburn，1974)，相应的方程即为伪得分方程或**估计方程**(estimating equations)。麦克莱(McCullagh，1983)指出伪似然估计量和最大似然估计量具有类似的属性，诸如由与最大似然相同的公式所给出的协方差矩阵，以及一致和渐近正态性。

### 6.8.2 迭代广义最小二乘

在残协方差矩阵 $\mathbf{V}_{(D)}$ 已知的多元线性模型中，参数可由广义最小二乘估计，其中，式(6.21)中的对角矩阵用(非对角的)协方差矩阵替代。由于残斜方差矩阵通常未知，我们必须使用迭代方法，如迭代广义最小二乘(IGLS)法。整个样本的响应向量的多层次线性混和模型可写为

$$\mathbf{y} = \mathbf{X}\beta + \boldsymbol{\Lambda}_{(D)}\boldsymbol{\zeta}_{(D)} + \boldsymbol{\epsilon}$$

令 $\mathbf{V}_{(D)}^k$ 为总残差 $\boldsymbol{\xi} = \boldsymbol{\Lambda}_{(D)}\boldsymbol{\zeta}_{(D)} + \boldsymbol{\epsilon}$ 的协方差矩阵的“当前”估计，回归参数可用 GLS 校正为

$$\beta^{k+1} = (\mathbf{X}'(\mathbf{V}_{(D)}^k)^{-1}\mathbf{X})^{-1}\mathbf{X}(\mathbf{V}_{(D)}^k)^{-1}\mathbf{y} \qquad (6.22)$$

用这些校正估计，给定一个协方差矩阵 $\mathrm{V}_{(\mathrm{D})}^{k+1}$ 的新的估计，方差参数 $\boldsymbol{\Psi}^{(l)}$ 可由残差

$\mathbf{r}^{k+1}=\mathbf{y}-\mathbf{X}'\boldsymbol{\beta}^{k+1}$来估计。具体来说，就是形成交叉乘积矩阵$\mathbf{r}^{k+1}\mathbf{r}^{k+1\prime}$且其期望等于$\mathbf{V}_{(D)}^{k+1}$。交叉乘积向量矩阵的期望可写成以方差参数作为系数的线性回归。例如，对于一个双层次随机截距模型(忽略上标)，则

$$\mathrm{E}[\mathrm{vec}(\mathbf{rr}')]=\mathrm{E}\begin{pmatrix} r_{11}^2 \\ r_{21}r_{11} \\ r_{12}^2 \\ r_{12}r_{11} \\ r_{21}r_{21} \\ \vdots \\ r_{n_jJ}^2 \end{pmatrix}=\begin{pmatrix} \psi+\theta \\ \psi \\ \psi+\theta \\ 0 \\ 0 \\ \vdots \\ \psi+\theta \end{pmatrix}=\psi\begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}+\theta\begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$$

方差和协方差参数可由广义最小二乘估计，其中 vec($\mathbf{rr}'$)的协方差矩阵可由前面的估计$\mathbf{V}_{(D)}^{k}$来推导。

IGLS 算法在模型的固定和随机部分的参数校正中迭代。所得到的估计是在$\boldsymbol{\zeta}_{(D)}$和$\boldsymbol{\epsilon}$的正态下的最大似然估计(Goldstein,1986)。戈德斯坦恩(Goldstein,1986)证明了非常高维的协方差矩阵$\mathbf{V}_{(D)}^{k}$的逆矩阵转换可通过其块对角结构来简化。

在收敛后，估计的回归参数的标准误可由上一个 GLS 步进，将协方差矩阵$\mathbf{V}_{(D)}$作为已知而估计得出。由于固定部分与随机部分的估计无关，这些标准误通常都是正确的。一个重要的例外是当响应为随机性缺失，且缺失依赖于观测响应，而非协变量时，如一个未包含协变量的纵贯性数据的线性随机截距模型。如果退出的概率随先于退出的观测响应的幅度增长而增长，一个大的随机截距方差(即高度组内相关)将意味着一个大的固定截距(因为对那些退出者的估算值会更高)，使得两个估计正相关。另请参见(Verbeke and Molenberghs,2000,第 21 章)。

关于 IGLS 算法的进一步细节，可见第 2 章和戈德斯坦恩(Goldstein,2003)中的附录 2.1。对 IGLS 进行一个小修正以约束迭代广义最小二乘(RIGLS)，将导致受约束最大似然(REML)估计(Goldstein,1989)。杨等人(Yang et al.,1999)为多层次结构方程模型提出了一种 IGLS 算法的扩展(另请参见 Rabe-Hesketh et al.,2001d)。

### 6.8.3 边缘和罚伪似然

边缘伪似然(MQL)和罚伪似然(PQL)已可由很多方法推导出来(见 6.3.1 节和 McCulloch and Searle,2001)。这里，我们根据戈德斯坦恩(Goldstein,2003)所给出的概述。

MQL 和 PQL 是基于由线性混和模型得到的近似广义线性混和模型，因此，IGLS 算法可被采用(**该算法不再对应于最大似然**)。该线性化方法类似于 6.8.1 节中介绍的迭代重加权最小二乘。

在广义线性混和模型中，响应的条件期望是(对于第一层次单元只保留下标 $i$)$\mu_i=h(\nu_i)$，其中 $h(\cdot)$是逆链接函数。在第 $k$ 次迭代的线性预测量的已知“当前”值处，通过将 $h(\nu_i)$扩展为一阶泰勒级数。

$y_i$ 的模型用 $h(\nu_i)$的一阶泰勒级数在第 $k$ 次迭代的线性预测的已知“当前”值处展开来线性化

$$\nu_i^k=\mathbf{x}_i'\boldsymbol{\beta}^k+\sum_{l=1}^{L}\mathbf{z}_i^{(l)\prime}\boldsymbol{\zeta}^{(l)k} \qquad (6.23)$$

得到

$$y_i \approx h(\nu_i^k) + \mathbf{x}_i'[\boldsymbol{\beta} - \boldsymbol{\beta}^k]h'(\nu_i^k) + \sum_l \mathbf{z}_i^{(l)}[\boldsymbol{\zeta}^{(l)} - \boldsymbol{\zeta}^{(l)k}]h'(\nu_i^k) + \epsilon_i \qquad (6.24)$$

其中，$\epsilon_i$ 是关于选定分布族的方差 $\phi V(\mu_i)$ 的非齐次误差项。请注意，这个表达式对于未知参数 $\boldsymbol{\beta}$ 是线性的。包含已知的当前参数 $\boldsymbol{\beta}^k$ 和 $\boldsymbol{\zeta}^{(l)k}$ 的项的和被当做偏移量 $o_i$，即

$$o_i = h(\nu_i^k) - h'(\nu_i^k)\mathbf{x}_i'\boldsymbol{\beta}^k - \sum_l h'(\nu_i^k)\mathbf{z}_i^{(l)}\boldsymbol{\zeta}^{(l)k}$$

而包含潜变量 $\zeta^{(l)}$ 的项对总残差 $\xi_i$ 有影响，在给定条件下，即

$$y_i = o_i + h'(\nu_i^k)\mathbf{x}'(\boldsymbol{\beta}) + \xi_i$$

则

$$\xi_i = \sum_l h'(\nu_i^k)\mathbf{z}_i^{(l)}\boldsymbol{\zeta}^{(l)} + \epsilon_i$$

将 $h'(\nu_i^k)$ 乘以 $\mathbf{X}_i$，用式(6.22)中的广义最小二乘，我们可得到 $\boldsymbol{\beta}^{k+1}$，其中 $V_{(D)}^k$ 是当前估计所有单元的总残差的协方差矩阵。

关于这个算法(Goldstein，1991，2003；Longford，1993，1994)有不同的方式。在边缘伪似然(MQL)中，式(6.23)和式(6.24)中的 $\boldsymbol{\zeta}^{(l)k}$ 设置为0。在罚伪似然(PQL)中，通过将潜变量设置为等于基于线性模型的后验众数的潜变量，$\boldsymbol{\zeta}^{(l)k} = \tilde{\boldsymbol{\zeta}}^{(l)k}$，从而改进展开。因而，MQL 和 PQL 的区别在于所用的偏移量。因为 MQL 将潜变量设定为0，故固定效应估计本质上是边缘效应。相对于需要的条件效应，这些差异减弱，正如4.8.1节所讨论的。

潜变量的估计量用“**工作变量**”更容易表示为

$$z_i = (y_i - o_i)/h'(\nu_i)$$

从而线性近似可用 GF 表达法(见4.2.2节)写为

$$\mathbf{z}_{(L)} = \mathbf{X}_{(L)}\boldsymbol{\beta} + \boldsymbol{\Lambda}_{(L)}\boldsymbol{\zeta}_{(L)} + g'(\mu_i)\boldsymbol{\epsilon}$$

并且潜变量可用下式校正

$$\boldsymbol{\zeta}_{(L)}^{k+1} = \boldsymbol{\Psi}_{(L)}\boldsymbol{\Lambda}_{(L)}'\mathbf{V}_{(L)}^{-1}(\mathbf{z}_{(L)} - \mathbf{X}_{(L)}\boldsymbol{\beta}^k)$$

其中，$\mathbf{V}_{(L)}$ 是某个顶层单元的 $\mathbf{V}_{(D)}$ 的一个块。随机部分的参数可由6.8.2节中的残差 $r_i = y_i - h(\nu_i)$ 来估计。因此，算法在给定 $\boldsymbol{\zeta}_{(L)}^k$ 和 $\boldsymbol{\Psi}^k$ 来校正 $\boldsymbol{\beta}$ 以及给定其他参数的当前值来校正 $\boldsymbol{\Psi}$ 和 $\boldsymbol{\zeta}_{(L)}$ 这两者之间迭代。

通过对潜变量用二阶泰勒级数展开，这个算法可得到显著的改进(Goldstein and Rasbash，1996)。对于 PQL，改进包括偏移量和总残差的方差。对于 MQL，偏移量不受影响，因为随机部分被设为零。更多的详情请见戈德斯坦恩一书中的附录4.1(Goldstein，2003)。

PQL 方法被 Goldstein(1991)、Schall(1991)、Breslow and Clayton(1993)、Longford(1993)、Wolfinger and O'Connell(1993)、Engel and Keen(1994)以及 McGilchrist(1994)等用于广义线性混和模型。由于避免了数值积分，因此，该算法在计算上非常有效。此外，该方法被用于包含交叉随机效应的模型(如 Breslow and Clayton，1993；以及用于线性混和模型，Goldstein，1987)，(空间域或时域)自相关随机效应(如 Breslow and Clayton，1993；Langford et al.，1999)，以及多重隶属模型(如 Hill and Goldstein，1998；Rasbash and Browne，2001)，另见3.2.6节。对于包含交叉随机效应的模型的另一种分析方法请见6.11.5节。然而，PQL 尚未用于因子负荷模型或结构方程模型中。

给定随机效应，当响应的条件分布接近于正态时，这些模型工作得很好，如均值大于或等于5的泊松分布(Breslow，2003)，或均值大于等于7的泊松分布(McCulloch and

Searle,2001),或响应与二项式的较大的分母成比例时。当属于同一个群组的响应的条件联合分布近似正态,或者说随机效应的后验分布近似正态时,这些模型也工作良好。即便对于二项响应,随着群组规模增加,也能有同样的效果。

然而,无论是 MQL 还是 PQL 对于群组规模小的二项响应效果都很差(如 Rodriguez and Goldman,1995,2001;Breslow and Lin,1995;Lin and Breslow,1996;Breslow et al.,1998;Goldstein and Rasbash,1996;Browne and Draper,2004;McCulloch and Searle,2001;Breslow,2003)。在这种情况下,PQL 的近似好于 MQL,并且随机部分的二阶展开(MQL-2 或 PQL-2)比一阶展开(MQL-1 或 PQL-1)能产生更好的结果。但是,罗德里克和戈德曼(Rodriquez and Goldman,2001)发现,对于“大的”随机效应方差,同时对固定和随机参数进行估计,比起 ML 或吉布斯抽样来,效果都会减弱,即便是用 PQL-2。他们指出,使用对于“小的”随机效应方差工作良好的方法是危险的,因为组内依赖程度几乎无法预先知道。不幸的是,MPL-2 和 PQL-2 有时数值不稳定,罗德里克和戈德曼(Rodriquez and Goldman,2001)报告了他们所遇到的一个问题。

$\hat{\boldsymbol{\beta}}$ 的标准误并没考虑 $\hat{\boldsymbol{\Psi}}$ 估计的不精确。这将会导致大的偏误,因为固定效应估计通常与广义线性混和模型中的方差估计相关。这在 PQL 中是显而易见的,因为用于估计固定效应的偏移量依赖于方差。边缘和罚伪似然的另一个缺点是无法提供似然,因而无法进行诸如似然比值检测,AIC,BIC 等模型选择标准,以及基于诊断的似然,等等。

## 6.9 广义估计方程(GEE)

使用“广义估计方程”(GEE)的估计最早由梁、泽戈及其同事(见 Liang and Zeger,1986;Zeger and Liang,1986;Zeger et al.,1988)在他们的一系列论文中所提出。GEE 可被看做是 6.8.1 节中所介绍的伪似然方法对多元回归模型的推广。对于依赖响应来说,比如重复测量或群组数据,该方法应用相当普遍。使用广义线性模型对响应的(关于任意潜变量的)边缘期望进行建模。对于群组 $j$ 包含 $n_j$ 个观测的双层次模型,则

$$g[\mathrm{E}(y_{ij} \mid \mathbf{x}_{ij})] = \mathbf{x}'_{ij}\boldsymbol{\beta}$$

回归参数 $\boldsymbol{\beta}$ 则表示**边缘**或**总体平均效应**。重要的是,这些效应与条件或群组特定效应不同(见图 1.6 及 3.6.5 节)。

现在我们来考虑给定协变量条件下,响应的边缘方差及协方差。方差被假设为 $\phi V(\mu_{ij})$,以对应于所设定的广义线性模型(见 2.2 节)。将这些方差与由 $\boldsymbol{\alpha}$ 所确定的**工作相关矩阵** $\mathbf{R}_j(\boldsymbol{\alpha})$ 合并,则协方差矩阵成为

$$\mathbf{V}_j = \mathbf{B}_j^{1/2}\mathbf{R}_j(\boldsymbol{\alpha})\mathbf{B}_j^{1/2}\phi$$

其中,$\mathbf{B}_j$ 为对角矩阵,其元素为 $b''(\theta_{ij}) = V(\mu_{ij})$。

式(6.18)中的伪得分方程现在即可推广为如下形式的**广义估计方程**

$$S_{\beta}(\boldsymbol{\beta},\boldsymbol{\alpha}) = \sum_j \frac{\partial \boldsymbol{\mu}'_j}{\partial \boldsymbol{\beta}}\mathbf{V}_j^{-1}[\mathbf{y}_j - \boldsymbol{\mu}_j] = \mathbf{0}$$

该方程不仅取决于边缘效应 $\beta$,还取决于依赖参数 $\boldsymbol{\alpha}$。在此

$$\frac{\partial \boldsymbol{\mu}_j}{\partial \boldsymbol{\beta}} = \mathbf{B}_j\boldsymbol{\Delta}_j\mathbf{X}_j$$

其中,$\boldsymbol{\Delta}_j$ 是元素为 $\frac{\partial \theta_{ij}}{\partial v_{ij}}$ 的对角矩阵。

梁和泽戈(Liang and Zeger,1986)提出在下面两个步骤中进行迭代:①(给定 $\boldsymbol{\alpha}$ 及

$\phi$)估计$\boldsymbol{\beta}$,求解广义估计方程;②(给定$\boldsymbol{\beta}$)基于皮尔逊残差$r_{ij}^{P}$,估计$\boldsymbol{\alpha}$及$\phi$。则

$$r_{ij}^{P} = [y_{ij} - b'(\hat{\theta}_{ij})]/[b''(\hat{\theta}_{ij})]^{\frac{1}{2}}$$

对于过度离散参数$\phi$,他们提出一个矩估计量,即

$$\hat{\phi} = \frac{1}{N}\sum_{j=1}^{N}\frac{1}{n_j}\sum_{i=1}^{n_j}(r_{ij}^{P})^2$$

对于不同的相关结构$\mathbf{R}_j(\boldsymbol{\alpha})$中的$\boldsymbol{\alpha}$,他们提出下述矩估计量:

• "独立的":

——相关结构:

$$\text{Cor}(y_{ij}, y_{i'j}) = 0$$

• "可换的":

——相关结构:

$$\text{Cor}(y_{ij}, y_{i'j}) = \alpha, \quad i \neq i'$$

——估计量:

$$\hat{\alpha} = \frac{1}{N}\sum\nolimits_{j=1}^{N}\frac{1}{n_j(n_j-1)}\sum\nolimits_{i\neq i'} r_{ij}^{P} r_{i'j}^{P}$$

• "AR(1)":

——相关结构:

$$\text{Cor}(y_{ij}, y_{i+t}, j) = \alpha^{t}, \quad t = 0, 1, \cdots, n_j - i$$

——估计量:

$$\hat{\alpha} = \frac{1}{N}\sum\nolimits_{j=1}^{N}\frac{1}{(n_j-1)}\sum\nolimits_{i\leqslant n_j-1} r_{ij}^{P} r_{i+1,j}^{P}$$

• "非结构化的":

——相关结构:

$$\text{Cor}(y_{ij}, y_{i'j}) = \alpha_{ii'}, \quad i \neq i'$$

——估计量:

$$\hat{\alpha}_{ii'} = \frac{1}{N}\sum\nolimits_{j=1}^{N} r_{ij}^{P} r_{i'j}^{P}$$

梁和泽戈(Liang and Zeger,1986)证明,随着群组个数增加,所估计的边缘效应$\hat{\boldsymbol{\beta}}$为渐近正态并且是一致的。重要的是,这些估计是"稳健"的,因为对于设定错误的相关结构,通过假设均值结构设定正确,则它们仍然是一致估计。所估计的边缘效应的协方差矩阵的一致估计,通常使用 8.3.3 节所介绍的夹挤估计量。

除了使用上面的$\alpha$的矩估计量,普伦迪斯(Prentice,1988)建议对相关参数$\alpha$(或者$\phi$)增加第二组估计方程。定义皮尔逊残差乘积向量为

$$\mathbf{u}_j = (r_{1j}^{P} r_{2j}^{P}, r_{1j}^{P} r_{3j}^{P}, \cdots, r_{n_j-1,j}^{P} r_{n_j,j}^{P})'$$

以及对角矩阵为

$$\mathbf{W}_j = \text{diag}\{\text{Var}(r_{1j}^{P} r_{2j}^{P}), \text{Var}(r_{1j}^{P} r_{2j}^{P}), \cdots, \text{Var}(r_{n_j-1,j}^{P} r_{n_j,j}^{P})\}$$

$\boldsymbol{\alpha}$的估计方程可表示为

$$S_{\alpha}(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \sum_{j}\frac{\partial \text{E}(\mathbf{u}_j)'}{\partial \boldsymbol{\alpha}}\mathbf{W}_j^{-1}[\mathbf{u}_j - \text{E}(\mathbf{u}_j)] = \mathbf{0}$$

对于类别响应来说,皮尔逊相关并不是合适的关联度的量度方法,这些基于皮尔逊相关的方法将存在问题。对于二项响应这样的特例来说,这些方法还能适用,但是仍然存在一个问题,即相关的容纳范围取决于边缘概率(如 Lord and Novick,1968;Bishop

et al.,1975)。对于类别数据的关联度通常采用优比,利普希茨等(Lipsitz et al.,1991)提出一个基于边缘优比的参数化方法。通常将优比进行结构化,以简化工作相关矩阵,如设定一个公共优比。对数线性模型还可设定优比依赖于协变量。一般来说,对工作相关矩阵的设定必然需要在简洁性和由于设定错误所带来的效率损失之间进行取舍(如Fitzmaurice,1995)。

赵和普伦迪斯(Zhao and Prentice,1990)以及梁等(Liang et al.,1992)提出将上述一阶估计方程(GEE-1)扩展为二阶估计方程(GEE-2)。在此,联合估计方程

$$S_{\beta,\alpha}(\boldsymbol{\beta},\boldsymbol{\alpha}) = \sum_j \frac{\partial[\boldsymbol{\mu}_j, \mathrm{E}(\mathbf{u}_j)]'}{\partial[\boldsymbol{\beta},\boldsymbol{\alpha}]}\begin{pmatrix} \mathbf{V}_j & \\ \mathrm{Cov}(\mathbf{u}_j,\mathbf{y}_j) & \mathbf{W}_j \end{pmatrix}^{-1}\begin{pmatrix} \mathbf{y}_j - \boldsymbol{\mu}_j \\ \mathbf{u}_j - \mathrm{E}(\mathbf{u}_j) \end{pmatrix} = \mathbf{0}$$

可同时对$\boldsymbol{\beta}$和$\boldsymbol{\alpha}$求解。GEE-2 相比 GEE-1 的主要优点在于效率上的提高,特别是对于$\alpha$。但是,却损失了 GEE-1 的稳健性,因为 GEE-2 取决于对依赖结构的正确设定。另外,对所需的估计 $\mathrm{Cov}(\mathbf{u}_j,\mathbf{y}_j)$ 也更加棘手。

当用边缘优比来表示依赖性时,凯里等(Carey et al.,1993)提出用包含一个偏移量的对数概率回归来估计$\alpha$。这种 GEE 实现方法被称为"交替对数概率回归(alternating logistic regression,ALR)"。ALR 保留了 GEE-1 对$\boldsymbol{\beta}$估计的稳健性,而同时它对$\boldsymbol{\alpha}$的估计与更复杂的 GEE-2 方法几乎同效。

由于显示避免了积分,对于纵贯性及群组数据模型的估计来说,GEE 方法毫无疑问是个重大贡献。在 9.2 节中,我们使用 GEE 来分析呼吸道感染的纵贯性数据,同时与随机效应模型进行比较。有趣的是,近来 GEE 被扩展到主要关注依赖结构的因子模型(Reboussin and Liang,1998)中。

GEE 方法的一个重要缺点在于对缺失数据的处理只能被严格假设为完全随机性缺失(MCAR),否则的话,估计方程将会偏误(如 Rotnitzky and Wypij,1994)。然而,不太被注意到的是,缺失可能实际上依赖于协变量,而非观测响应(Little,1995)。罗宾斯等(Robins et al.,1994)提出将估计方程与逆概率加权合并,在正确设定缺失机制的情况下,可得到一致估计量。

GEE 方法的另外一个缺点在于,一般来说难以评估模型的合适性(如 Albert,1999),如不存在基于似然的诊断方法。GEE 应只用于关于边缘或总体平均效应的研究,而避免用于病因学分析。这是因为因果过程在群组或个体的层次上起作用,而非在总体层次上。因而,总体平均效应只是描述,并很大程度上取决于总体的异质性程度。最后,林赛和兰伯特(Lindsey and Lambert,1998)以及克劳奇莱和戴维斯(Crouchley and Davies,1999)指出,如果存在内生协变量,如纵贯性数据中的"基准"(初始)响应,则所估计的回归参数就不再是一致估计。

## 6.10 固定效应方法

本书的重点,包括在本章中前面的讨论,集中在潜变量作为随机变量的情况。在这一节中,我们暂时偏离一下,将潜变量作为**未知固定参数**。

### 6.10.1 联合最大似然

乍一看,可能会很自然地试图去同时估计基本参数 $\boldsymbol{\vartheta}$ 和"潜得分",或在给定潜变量条件下,通过最大化响应的似然所得到的潜变量 $\boldsymbol{\zeta}$ 的数值。对于双层次随机截距模型,对数似然

$$\ln \prod_{ij} g^{(1)}(y_{ij} \mid \mathbf{x}_{ij}, \zeta_j; \boldsymbol{\vartheta})$$

为同时关于 $\boldsymbol{\beta}$ 和 $\zeta_j, j=1,\cdots,J$ 的最大化。因此,随机截距只是简单地被当做为固定参数并跟 $\boldsymbol{\beta}$ 一起估计。应该注意的是,此处的似然不同于6.3.1节式(6.7)中的 $h$-似然,$h$-似然是响应和潜变量的联合似然,即同时关于参数和潜得分的联合最大化。

在三层次随机截距模型中,对第三层次截距的估计需要对第二层次截距进行约束,如将每个第三层次群组的截距之和约束为零。因此,忽略所有含有对应较低层次潜变量的高层次潜变量就显得更自然。因此,我们假定该模型仅包括第二层次潜变量。

待估计的潜得分的数量随着第二层次群组 $J$ 的数量的增加而增加,这可能带来一个根本问题。这就是众所周知的在存在"附加参数"的情况下估计"结构参数"的问题。在本书中,基本参数被视为结构的,而潜得分则被视为附加的。基本的问题是,在"附加参数"存在的情况下,结构参数的最大似然估计量不一定具有一致性(Neyman and Scott, 1948;又见 Lancaster,2000)。泽尔纳(Zellner,1971:114-115)用一个简单的例子证明了这种不一致性是如何产生的。

更重要的是,对于包含正态分布残差的线性模型或对泊松分布残差的对数线性模型来说,是不存在附加参数问题的。在这些模型中,估计与将在6.10.2节中所介绍的条件最大似然相一致。进一步讨论详见卡梅隆和特里维迪(Cameron and Trivedi,1998)。

对于包含连续响应的公共因子模型,安德森和鲁宾(Anderson and Rubin,1956)证明了劳利(Lawley,1942)所提出的似然并不存在最大值。麦当劳(McDonald,1979)建议使用**最大似然比值**(maximum likelihood ratio,MLR)估计量作为解决这一问题的方法,其中可将 $\boldsymbol{\Psi}$ 当做任一正定矩阵。事实证明,由此所产生的结构参数估计与在随机因子情形下用ML所获得的估计相等,然而,给定"不确定"因子得分的科斯托曼表达式(Kestelman, 1952,Guttmann,1955),对因子得分的估计却不一致。

现在我们转向二项响应,安德森(Andersen,1973)设定了一个具有群组特定截距的对数回归模型的分支模型,其中,包含 $n=2$ 个第一层次单元以及一个二分协变量(又见 Chamberlain,1980 以及 Breslow and Day,1980)。具体地说,安德森证明,对于 $J\rightarrow\infty$,联合最大似然估计量是不一致的,因为 $\hat{\boldsymbol{\beta}}$ 的概率收敛为 $2\boldsymbol{\beta}$。由卡兹(Katz,2001)所做的模拟结果表明,联合似然在 $n_j=n>20$ 的情况下是安全的,在 $8<n<16$ 时,也许是可以接受的。对于拉什模型,模拟表明,$\hat{\boldsymbol{\beta}}$ 收敛于 $\frac{n}{n-1}\boldsymbol{\beta}$,其中,$n$ 是题器个数。哈伯曼(Haberman, 1977)证明,这种偏误将在 $J\rightarrow\infty$,$n\rightarrow\infty$ 和 $\frac{J}{n}\rightarrow\infty$ 时消失。赫克曼(Heckman,1981b)对于固定效应正态概率回归进行了蒙特卡洛实验,并发现少量偏误,且总是接近为零。

在二项响应情况下数值问题非常常见的。如果对一个群组的所有响应都是零(或1),联合最大似然估计 $\hat{\zeta}_j$ 将偏向为 $-\infty$($\infty$)。此外,如果拉什模型中有一个题器 $i$ 都失败(或通过),则 $\hat{\beta}_i$ 将趋于 $-\infty$($\infty$)。请注意,后者也是边缘最大似然估计中的一个问题。对于双参数对数概率IRT模型,莱特(Wright,1977)发现,为防止估计出现偏向,对判别参数通常必须施加一个任意给定的上限约束。联合最大似然的另一个重要的不足是无法包括群组层次的协变量。

## 6.10.2 条件最大似然

对于3.3.4节所讨论的单参数对数概率IRT模型或拉什模型,拉什(Rasch,1960)建

议使用条件最大似然估计,即

$$\Pr(y_{ij}=1)=\frac{\exp(\beta_i+\zeta_j)}{1+\exp(\beta_i+\zeta_j)}$$

更重要的是,这一方法避免了附加参数问题。

条件最大似然方法要求存在参数 $\zeta_j$ 的充分统计。然后似然在这些充分统计的条件下最大化。

对具有固定的群组特定截距 $\zeta_j$ 的对数概率回归模型,线性预测量则为 $\nu_{ij}=x'_{ij}\beta+\zeta_j$。对于二项响应,$y_{ij}=0,1$,该模型则为

$$\Pr(y_{ij}=1\mid \mathbf{x}_{ij})=\frac{\exp(\mathbf{x}'_{ij}\boldsymbol{\beta}+\zeta_j)}{1+\exp(\mathbf{x}'_{ij}\boldsymbol{\beta}+\zeta_j)} \qquad (6.25)$$

拉什模型即为当 $\mathbf{x}'_{ij}\boldsymbol{\beta}=\beta_i$ 时,固定效应对数概率回归的一种特殊情形。

由于群组特定截距,在群组 $j$ 内的所有响应是独立分布的,其联合概率为

$$\Pr(\mathbf{y}_j)=\prod_{i=1}^{n_j}\Pr(y_{ij}=1)^{y_{ij}}[1-\Pr(y_{ij}=1)]^{1-y_{ij}}$$

将式(6.25)代入 $\Pr(y_{ij}=1)$ 并重写方程,则联合概率可写为

$$\Pr(y_j)=\exp\left[\zeta_j\sum_{i=1}^{n_j}y_{ij}+\boldsymbol{\beta}'\sum_{i=1}^{n_j}\mathbf{x}_{ij}y_{ij}+a(\zeta_j,\boldsymbol{\beta})\right]$$

其中

$$a(\zeta_j,\boldsymbol{\beta})=\ln\prod_{i=1}^{n_j}[1+\exp(\mathbf{x}'_{ij}\boldsymbol{\beta}+\zeta_j)]$$

由指数族分布理论,$\sum_{i=1}^{n_j}y_{ij}$是群组特定参数 $\zeta_j$ 的一个最小充分统计。从而,条件分布 $\Pr(\mathbf{y}_j\mid\sum_{i=1}^{n_j}y_{ij})$ 不依赖于 $\zeta_j$。条件最大似然的理念是通过最大化不包含"附加参数"$\zeta_j$ 的 $\prod_j\Pr(y_j\mid\sum_{i=1}^{n_j}y_{ij})$ 来估计"结构参数"。

联合概率也可表示为

$$\Pr(\mathbf{y}_j)=\frac{\exp\left[\zeta_j\sum_{i=1}^{n_j}y_{ij}+\boldsymbol{\beta}'\sum_{i=1}^{n_j}\mathbf{x}_{ij}y_{ij}\right]}{\prod_{i=1}^{n_j}\left[1+\exp(\mathbf{x}'_{ij}\boldsymbol{\beta}+\zeta_j)\right]}$$

而 $\tau_j=\sum_{i=1}^{n_j}y_{ij}$ 的概率可表示为

$$\Pr\left(\sum_{i=1}^{n_j}y_{ij}=\tau_j\right)=\sum_{\mathbf{d}_j\in B(\tau_j)}\frac{\exp\left[\zeta_j\tau_j+\boldsymbol{\beta}'\sum_{i=1}^{n_j}\mathbf{x}_{ij}d_{ij}\right]}{\prod_{i=1}^{n_j}\left[1+\exp(\mathbf{x}'_{ij}\boldsymbol{\beta}+\zeta_j)\right]},\tau_j=0,1,\cdots,n_j$$

其中,$B(\tau_j)=\{\mathbf{d}_j=(d_{1j},\cdots,d_{n_j,j}):d_{ij}=0\ or\ 1,\ \sum_{i=1}^{n_j}d_{ij}=\tau_j\}$,$B(\tau_j)$ 中元素的个数为 $\binom{n_j}{\tau_j}$,并随群组规模的增加而迅速增加。现在通过将 $\Pr(\mathbf{y}_j)$ 除以 $\Pr(\sum_{i=1}^{n_j}y_{ij}=\tau_j)$,可得到条件概率 $\Pr(\mathbf{y}_j\mid\sum_{i=1}^{n_j}y_{ij}=\tau_j)$,即

$$\Pr\left(\mathbf{y}_j\,\middle|\,\sum_{i=1}^{n_j}y_{ij}=\tau_j\right)=\frac{\exp\left[\boldsymbol{\beta}'\sum_{i=1}^{n_j}\mathbf{x}_{ij}y_{ij}\right]}{\sum_{\mathbf{d}_j\in B(\tau_j)}\exp\left[\boldsymbol{\beta}'\sum_{i=1}^{n_j}\mathbf{x}_{ij}d_{ij}\right]}$$

值得一提的是,当 $\tau_j=1$ 时,该概率具有与多项式对数概率模型相同的形式(见2.4.3节),对于考克斯回归,该概率具有偏似然贡献(见2.5节)。

条件似然 $l_C(\beta | \mathbf{X})$ 则变为

$$l_C(\boldsymbol{\beta} \mid \mathbf{X}) = \prod_{j=1}^{N} \frac{\exp\left[\boldsymbol{\beta}' \sum_{i=1}^{n_j} \mathbf{x}_{ij} y_{ij}\right]}{\sum_{\mathbf{d}_j \in B(\tau_j)} \exp\left[\boldsymbol{\beta}' \sum_{i=1}^{n_j} \mathbf{x}_{ij} d_{ij}\right]}$$

该条件似然并不依赖于附加参数 $\zeta_j$。需要注意的是,包含 $\tau_j=0$ 或 $\tau_j=n_j$ 的群组,仅有零响应或单元响应,对似然并没有贡献,因为它们的条件概率变成 1。因此,很可能出现数据的大量损失,特别是当群组规模较小时。

对于群组规模为 2,也就是当 $n_j=2$ 这一特殊情形,分析其条件似然会很有指导意义。在这里,唯一有贡献的信息是 $\tau_j=1$,即 $(y_{1j}=0, y_{2j}=1)$ 和 $(y_{1j}=1, y_{2j}=0)$。前者的条件概率则成为

$$\Pr(y_{1j}=0, y_{2j}=1 \mid y_{1j}+y_{2j}=1) = \frac{\Pr(y_{1j}=0, y_{2j}=1)}{\Pr(y_{1j}=0, y_{2j}=1) + \Pr(y_{1j}=1, y_{2j}=0)}$$
$$= \frac{\exp[\boldsymbol{\beta}'(\mathbf{x}_{2j}-\mathbf{x}_{1j})]}{1+\exp[\boldsymbol{\beta}'(\mathbf{x}_{2j}-\mathbf{x}_{1j})]}$$

由此产生的条件似然就简化为协变量为 $\mathbf{X}_{2j}-\mathbf{X}_{1j}$ 的二项响应的一个(非条件)对数概率似然,当(($y_{1j}=0, y_{2j}=1$),二项响应 $y'_j=1$;当($y_{1j}=0, y_{2j}=1$))时,二项响应 $y'_j=0$(忽略一致响应)。重要的是,单元之间没有变化的且与协变量有关的 $\boldsymbol{\beta}$ 的元素不能通过条件最大似然来估计,因为与 $\mathbf{x}_{2j}-\mathbf{x}_{1j}$ 相对应的元素对于所有的 $j$ 都为零。在(配对的)对照研究中,响应为($y_{1j}=0, y_{2j}=1$),$y_{2j}$ 和 $y_{1j}$ 分别表示病例组和控制组。根据上述结果,对于包含常数响应 $y'_j=1$ 和协变量 $\mathbf{x}_{2j}-\mathbf{x}_{1j}$ 的对数概率模型,条件似然就简化为(非条件)似然。另见霍尔福德等(Holford et al.,1978)。

安德森(Andersen,1970)证明,在较弱的正则条件下,条件最大似然可得到一致的且渐近的正态估计量。对于指数族,他还证明,在"*S* **辅助**(S-ancillarity)"下,估计是渐近有效的(如 Barndorff-Nielsen,1978)。安德森(Andersen,1973)证明,这种情形对拉什模型也是成立的。

条件最大似然的一个优势在于,我们无须对潜变量的分布做出任何假设。重要的是,这意味着,基于这个方法的估计比基于随机效应的可能更加"稳健"。如果随机效应模型是正确的,则使用条件最大似然估计将有效率上的损失,但安德森(Andersen,1973)证明,这个损失是较小的。另一个重要的优点是,潜变量与协变量之间的相关将不会成为问题,这一特点在计量经济学学中受到了大量的关注(如 Hausman,1978)。相比之下,随机效应模型通常假定随机效应和协变量无关。然而,在随机效应模型中,这一假设对有些协变量可以放宽,可通过将这些协变量的群组均值作为额外估计量包含到模型中来实现,见 3.2.1 节。

不幸的是,依靠条件最大似然严重限制了可估计的模型类型。构建一个条件似然估计所需要的充分统计只在简单的具有群组特定截距的模型中存在。此外,它要求这些模型属于指数族并具有典型的链接函数。对于连续响应,这种联系是标准的同一性链接;对于二项响应则为对数概率连接(与正态概率相对照);对于计数则为对数链接;而对无序多分类的响应则为多项式对数概率链接(如 Andersen,1973;Chamberlain,1980)。如果包括群组特定的协变量,即使对具有群组特定截距的简单模型也无法估计。如 3.6.1 节所讨论的,对组内协变量的回归参数的估计,仅组内效应。相比之下,随机效应估计则是对组内效应和组间效应的加权平均(在平衡数据的情形下,等于组内效应)。与随机效应模型比较的另一个缺点就是,我们无法得到一个群组特定的固定效应对协变量或其他群

组特定效应回归的结构模型。

## 6.11 贝叶斯方法

### 6.11.1 简 介

在贝叶斯方法中,没有潜变量和参数的区别;它们都被认为是随机量。设 D 表示观测数据,L 表示参数以及潜变量(包括缺失数据)。推断要求对所有随机量建立一个联合概率分布 Pr(**D**,**L**)。联合分布由两部分组成:**似然** Pr(**D**|**L**)和**先验分布** Pr(**L**)。设定 Pr(**D**|**L**)和 Pr(**L**),即给出了**全概率模型**(full probability model),其中

$$\Pr(\mathbf{D},\mathbf{L}) = \Pr(\mathbf{D}\mid \mathbf{L})\Pr(\mathbf{L})$$

在观测到 D 之后,根据贝叶斯定理可得到在给定 **D** 后的 **L** 的**后验分布**为

$$\Pr(\mathbf{L}\mid \mathbf{D}) = \frac{\Pr(\mathbf{D}\mid \mathbf{L})\Pr(\mathbf{L})}{\int \Pr(\mathbf{D}\mid \mathbf{L})\Pr(\mathbf{L})d\mathbf{L}}$$

不严格地来讲,后验分布用观测数据(由似然所代表)的信息来校正先验的“知识”(由先验分布所代表)。需要注意的是,后验分布是与先验分布和似然的乘积成比例的。

所有关于未知量的信息都包含在给定数据 **D** 的后验分布中。然而,要理解可能有数千维的后验分布是很难的甚至是不可能的,因而,贝叶斯学派通常概述在后验中的信息。一种首选的概述即是参数的后验期望,因为这些“估计”使二次损失函数(估计的均方差)的后验期望最小化。期望的另一优点,就是它对一个参数子集的值不随对其余的参数的边缘化的变化而变化。与此相反,后验众数在边缘化下并非守恒。后验的其他特性也用于贝叶斯推断,其中包括矩、分位数和最高后验密度域。所有这些量都可表示为 **L** 函数的后验期望。

### 6.11.2 贝叶斯众数或众数后验(MAP)

在马尔可夫链蒙特卡洛(MCMC)方法出现之前(见下面),贝叶斯众数或众数后验(MAP)法,通常被用来逼近期望,因为众数往往容易被数值近似。林赛和斯密斯(Lindley and Smith,1972)建议,推断可基于**联合**后验众数,通过使用标准优化方法可得到此联合后验众数。然而,对参数子集的后验期望用边缘化其他参数后的众数通常会有更好的逼近(如 O'Hagan,1976)。关于结构参数 $\theta$ 的推断,最好的办法是考虑附加参数和对潜变量积分后的边缘后验的众数。使用联合后验众数的另一个问题就是附加参数的问题可能会出现,从而使贝叶斯众数预测的渐近正态性失效。因此,必须谨慎行使,以确保贝叶斯众数预测与在大样本情况下的贝叶斯(均值)预测相吻合。在 IRT 模型的范畴中,联合后验众数已经被斯瓦米纳坦和吉福德(Swaminathan and Gifford,1982,1985,1986)用于单参数、双参数及三参数的对数概率模型中。

### 6.11.3 多层贝叶斯模型

潜变量模型可以被看做是**多层**贝叶斯模型。这是因为一些参数(如潜变量)的先验分布取决于更多的参数(潜变量的方差和协方差),即所谓的**超参数**(hyperparameters)。这些超参数的分布被称为**超先验**。在第 1 阶段,响应的分布 Pr($\mathbf{D}|\boldsymbol{\zeta},\boldsymbol{\vartheta}_1$)被设定为依赖于潜变量 $\boldsymbol{\zeta}$ 和参数 $\boldsymbol{\vartheta}_1$。在第二阶段,设定参数的先验分布 Pr($\boldsymbol{\vartheta}_1$)和潜变量的先验分布

$\Pr(\boldsymbol{\zeta}|\boldsymbol{\vartheta}_2)$，后者依赖于超参数 $\boldsymbol{\vartheta}_2$。在第3阶段，为超参数设定一个超先验 $\Pr(\boldsymbol{\vartheta}_2)$。在广义线性混合模型中，$\boldsymbol{\vartheta}_1$ 为回归，也可能是离散和阈值参数，而 $\boldsymbol{\vartheta}_2$ 则是随机效应的方差和协方差参数。给定数据下参数的后验分布可写为

$$\Pr(\boldsymbol{\zeta},\boldsymbol{\vartheta}_1,\boldsymbol{\vartheta}_2 \mid \mathbf{D}) \propto \Pr(\mathbf{D} \mid \boldsymbol{\zeta},\boldsymbol{\vartheta}_1)\Pr(\boldsymbol{\vartheta}_1)\Pr(\boldsymbol{\zeta} \mid \boldsymbol{\vartheta}_2)\Pr(\boldsymbol{\vartheta}_2) \qquad (6.26)$$

因此，这里"多层"并非指数据结构，而是指设定模型的次序。林赛等(Lindley et al., 1971)，林赛和斯密斯(Lindley and Smith, 1972)以及斯密斯(Smith, 1973)给出了贝叶斯多层线性模型，并被诺维卡等(Novick et al., 1973)应用。"频率学派"的潜变量模型可被视为"经验贝叶斯"，这是因为潜变量"先验"分布的参数(方差和协方差)是估计出来的而非假设一个"超先验"的分布。

### 6.11.4 先验分布

用先验分布(和超分布)似乎有4种不同的动机：

第一，是"真正"的贝叶斯方法，不同于其他几种是"实用"的贝叶斯方法。真正的贝叶斯事先设定参数的反映先验信知或已知信息的先验分布。如，在独立群组因子模型中，因子负荷期望为正数，这种先验信知可通过适当的先验分布来表示。又如，一个在专家意见的基础上对临床试验治疗效应的先验估计(如 Spiegelhalter et al., 1994)。

第二，先验可用以确保估计限于一个允许的参数空间，如用以避免当独特因子具有负方差时的"海伍德(Heywood)特例"(如 Martin and McDonald, 1975)。在潜类模型中，先验有时是用来防止当条件响应概率接近零或一，而其相应的对数概率参数接近 $-\infty$ 和 $\infty$ 时的边界解。

第三，先验有助于模型辨识。例如，可设定一个测量误差方差先验来辨识一个没有包含重复测量的测量模型。与将方差视为已知参数的常规方法相比，这种做法可能是更可取的，因为先验能反映参数的不确定性。同样的，先验可用于解决在三参数对数概率 IRT 模型中标准误差过多这一问题(Wainer and Thissen, 1982)。

最后，也许是先验的最实用的地方在于，马尔可夫链蒙特卡洛(MCMC)方法(见下文)就可用于估计复杂的模型，而对于复杂模型其他方法表现得很坏或不适用，如对包含交叉随机效应的广义线性混合模型。先验通常被设定为"无信息的"(也指平的，模糊或分散)，以尽量减少其对统计推断的影响。然而后验中的似然分量占主导地位，因此，后验几乎与似然成比例。这种实用方法也反映在有关贝叶斯模型的现代书籍中，如康登(Congdon, 2001)和流行的 BUGS 软件中所附的例子(Spiegelhalter et al., 1996bc)。

### 6.11.5 马尔可夫链蒙特卡洛

MCMC 方法的目的是从后验分布中抽取出参数和潜变量 $\boldsymbol{\varphi}=\mathbf{L}$，用以获得函数 $f(\boldsymbol{\varphi})$ 的期望。如同在6.3.3节描述的原始蒙特卡洛积分，从 $h(\boldsymbol{\varphi})$ 中抽取**独立**样本可能是行之不通的。然而，对期望 $\mathrm{E}[f(\boldsymbol{\varphi})]$ 的一致估计量可从**依赖**样本中获得，只要样本是按正确的比例完全从 $h(\boldsymbol{\varphi})$ 中抽取的。这个可以通过用具有目标分布 $h(\boldsymbol{\varphi})$ 的马尔可夫链作为其稳态分布，从而导出马尔可夫链蒙特卡洛(MCMC)。

设($\boldsymbol{\varphi}^{(0)},\boldsymbol{\varphi}^{(1)},\cdots$)为一个随机变量序列。在**一阶同质马尔可夫链**(first order homogenous Markov chain)中，下一状态 $\boldsymbol{\varphi}^{(r+1)}$ 是从分布 $P(\boldsymbol{\varphi}^{(r+1)}|\boldsymbol{\varphi}^{(r)})$ 中抽取，并仅依赖于当前状态 $\boldsymbol{\varphi}^{(r)}$，而非依赖于链 $\{\boldsymbol{\varphi}^{(0)},\boldsymbol{\varphi}^{(1)},\cdots,\boldsymbol{\varphi}^{(r-1)}\}$ 的"历史"或 $r$。更重要的是，链将逐步"忘记"它的初始状态，并最终收敛于一个特定的稳态分布。为了获得所需的分布，我们忽略那些在我们认为已经达到平稳之前的所有状态，并将其被称为"老化"期。有时

使用具有不同初始状态的几个链来监测收敛直到一个稳态分布。过了老化期后，对$f(\boldsymbol{\varphi})$样本平均数我们需要运行足够长时间的链，以获得可靠的期望估计；而要确定运行多久并不容易，因为抽取是相互依赖的。

有多种方法可以将具有目标分布$h(\boldsymbol{\varphi})$的马尔可夫链建构为稳态分布。我们从最复杂的算法开始，并逐步介绍更简单的算法。

*Metropolis-Hastings* **算法**

由$\boldsymbol{\varphi}^{(r)}$，用下列步骤可得出$\boldsymbol{\varphi}^{(r+1)}$：

第1步：从提议分布$q(\check{\boldsymbol{\varphi}}|\boldsymbol{\varphi}^{(r)})$中选取一候选点$\check{\boldsymbol{\varphi}}$。如$q(\check{\boldsymbol{\varphi}}|\boldsymbol{\varphi}^{(r)})$可以是具有均值$\boldsymbol{\varphi}^{(r)}$和固定协方差矩阵的多元正态分布。

第2步：接受概率为

$$\alpha(\boldsymbol{\varphi}^{(r)},\check{\boldsymbol{\varphi}}) = \min\left(1,\frac{h(\check{\boldsymbol{\varphi}})q(\boldsymbol{\varphi}^{(r)}\mid\check{\boldsymbol{\varphi}})}{h(\boldsymbol{\varphi}^{(r)})q(\check{\boldsymbol{\varphi}}\mid\boldsymbol{\varphi}^{(r)})}\right)$$

的候选点。

一般来说，该算法步骤如下：在第1步，我们从一个方便的但不正确的分布中抽样。在第2步，我们通过一个正确的但是不太直观的方法来修正它。

如果候选点被接受，下一状态就变为$\boldsymbol{\varphi}^{(r+1)}=\check{\boldsymbol{\varphi}}$；如果候选点被否决，链便不会移动且$\boldsymbol{\varphi}^{(r+1)}=\boldsymbol{\varphi}^{(r)}$。因此，使用Metropolis-Hastings算法需要从提议分布中抽样并且计算用于获得接受概率的分数。更重要的，目标分布无须正态化，因此可以忽略后验分布的分母。

无论使用何种提议分布$q(\check{\boldsymbol{\varphi}}|\boldsymbol{\varphi}^{(r)})$，稳态分布都是$h(\boldsymbol{\varphi})$。然而，收敛速度当然将取决于提议分布与目标分布的接近程度。

*Metropolis* **算法**

该算法是Metropolis-Hastings算法的特例，它只考虑$q(\check{\boldsymbol{\varphi}}|\boldsymbol{\varphi}^{(r)})=q(\boldsymbol{\varphi}^{(r)}|\check{\boldsymbol{\varphi}})$的对称提议分布，如多元正态或$t$分布。在这种情况下，接受概率简化为

$$\alpha(\boldsymbol{\varphi}^{(r)},\check{\boldsymbol{\varphi}}) = \min\left(1,\frac{h(\check{\boldsymbol{\varphi}})}{h(\boldsymbol{\varphi}^{(r)})}\right)$$

这并不取决于提议分布。当$q(\check{\boldsymbol{\varphi}}|\boldsymbol{\varphi}^{(r)})=q(|\check{\boldsymbol{\varphi}}-\boldsymbol{\varphi}^{(r)}|)$时，则为**随机移动** Metropolis (random walk Metropolis)算法。

**单分量** *Metropolis-Hastings* **算法**

现在考虑将向量$\boldsymbol{\varphi}$分成$k$个分量，它可以是多个块但并不一定是标量，$\boldsymbol{\varphi}=\{\varphi_1,\varphi_2,\cdots,\varphi_k\}$。用单分量方法对$k$个分量每次更新一个$\varphi_i$，而无须对整个向量$\boldsymbol{\varphi}$进行更新。设$\varphi_i^{(r)}$表示迭代$r$结束时分量$\varphi_i$的状态。在完成迭代$r+1$的第$i-1$步后，定义$\boldsymbol{\varphi}_{-i}^{(r)}=\{\varphi_1^{(r+1)},\cdots,\varphi_{i-1}^{(r+1)},\varphi_{i+1}^{(r)},\cdots,\varphi_k^{(r)}\}$为不包含第$i$个元素的$\boldsymbol{\varphi}$。尽管可能有不同类型的随机顺序，我们仍然假定了一个固定的**更新次序**(updating order)。这样可能有助于对高度依赖性的分量进行比其他分量更频繁的校正(如Zeger and Karim, 1991)。

在单分量Metropolis-Hastings算法中，候选点$\check{\varphi}_i$取自于提议分布$q(\check{\varphi}_i|\varphi_i^{(r)},\boldsymbol{\varphi}_{-i}^{(r)})$。该候选点的接收概率为

$$\alpha(\boldsymbol{\varphi}_{-i}^{(r)},\varphi_i^{(r)},\check{\varphi}_i) = \min\left(1,\frac{h(\check{\varphi}_i\mid\boldsymbol{\varphi}_{-i}^{(r)})q(\varphi_i^{(r)}\mid\check{\varphi}_i,\boldsymbol{\varphi}_{-i}^{(r)})}{h(\varphi_i^{(r)}\mid\boldsymbol{\varphi}_{-i}^{(r)})q(\check{\varphi}_i\mid\varphi_i^{(r)},\boldsymbol{\varphi}_{-i}^{(r)})}\right)$$

其中，$h(\varphi_i|\boldsymbol{\varphi}_{-i})$是$\varphi_i$在$h(\boldsymbol{\varphi})$下的**完全条件分布**(full conditional distribution)，$\boldsymbol{\varphi}$的第$i$个分量的分布条件依赖于所有其他分量。更重要的是，当$h(\boldsymbol{\varphi})$源于条件独立模型时，$\alpha(\boldsymbol{\varphi}_{-i},\varphi_i,\check{\varphi}_i)$可简化。

**吉布斯抽样器**

吉布斯抽样器的基本理念是利用这样一个事实，即尽管在一个复杂的联合分布中，随机变量（块）的条件分布也有可能是相对简单的。这也是单分量 Metropolis-Hastings 算法中的一个特例，即用以校正 $\boldsymbol{\varphi}$ 的第 $i$ 个分量的提议分布 $q(\check{\varphi}_i|\varphi_i^{(r)},\boldsymbol{\varphi}_{-i}^{(r)})$ 是完全条件分布 $h(\check{\varphi}_i|\boldsymbol{\varphi}_{-i}^{(r)})$。值得注意的是，在 $\alpha(\boldsymbol{\varphi}_{-i}^{(r)},\varphi_i^{(r)},\check{\varphi}_i)$ 中，用 $q(\check{\varphi}_i|\varphi_i^{(r)},\boldsymbol{\varphi}_{-i}^{(r)})$ 代替 $h(\check{\varphi}_i|\boldsymbol{\varphi}_{-i}^{(r)})$，可得到一个接受概率

$$\alpha(\boldsymbol{\varphi}_{-i}^{(r)},\varphi_i^{(r)},\check{\varphi}_i) = 1$$

因此，在吉布斯抽样中，候选点总是被接受的。进而，目标分布是通过对向量 $\boldsymbol{\varphi}$ 执行随机移动来模拟，每次改变它的一个分量。最优的吉布斯抽样器就是当目标分布中 $\boldsymbol{\varphi}$ 的分量是独立的，这种情况下，每次迭代产生一个新的独立 $\boldsymbol{\varphi}$ 抽取。如果分量是高度相关的，则可通过对分量的正交化来提高收敛。

直接利用吉布斯抽样器需要样本可从来自目标分布的完全条件分布中抽取。当这不可行时，可利用更复杂的 Metropolis-Hastings 算法。另外，对常见的一元及**对数凹性**（log-concave）完全条件分布，吉尔克思和怀尔德（Gilks and Wild，1992）建议使用**自适应拒斥抽样**（Adaptive rejection sampling）（另见 Dellaportas and Smith，1993）。BUGS 软件（Spiegelhalter et al.，1996）也实现了该方法。对难以模拟的连续条件分布，阿尔伯特和奇布（Albert and Chib，1993）建议根据离散型来模拟。

例：随机截距正态概率模型的吉布斯抽样

考虑随机截距正态概率模型

$$\Pr(y_{ij}=1\mid \mathbf{x}_{ij}) = \Phi(\mathbf{x}_{ij}'\boldsymbol{\beta}+\zeta_j)$$

其中，$\zeta_j \sim \mathrm{N}(0,\psi)$。将此模型表达为潜响应模型

$$y_{ij}^* = \mathbf{x}_{ij}'\boldsymbol{\beta}+\zeta_j+\epsilon_{ij}$$

其中，$\epsilon_{ij}\sim\mathrm{N}(0,1)$，且独立于 $\zeta_j$。潜响应通过阈值函数 $y_{ij}$ 与观测响应相关联

$$y_{ij} = \begin{cases} 1 & 当\ y_{ij}^* > 0 \\ 0 & 当\ y_{ij}^* \leqslant 0 \end{cases}$$

对于回归系数 $\boldsymbol{\beta}$，我们考虑共轭多元正态先验

$$\boldsymbol{\beta} \sim \mathrm{N}_p(\boldsymbol{\beta}_0,\boldsymbol{\Sigma}_{\boldsymbol{\beta}})$$

并假定 $\boldsymbol{\beta}_0$ 与 $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ 已知。$\psi_1$ 的共轭逆伽玛（IG）密度被设定为

$$\Pr(\psi) = \frac{b^a}{\Gamma(a)}\psi^{-(a+1)}\exp^{-\frac{b}{\psi}},\psi>0$$

其中，$a>0$ 和 $b>0$ 分别为已知的形状和尺度参数，$\Gamma(\cdot)$ 是伽玛函数。假定 $\psi$ 与 $\boldsymbol{\beta}$ 相独立。后验分布 $\Pr(\mathbf{y}^*,\boldsymbol{\zeta},\psi,\beta|\mathbf{y},\mathbf{X})$ 是给定数据 $\mathbf{X}$ 和 $\mathbf{y}$ 条件下，未知参数（$\boldsymbol{\beta}$ 和 $\psi$）、潜变量 $\boldsymbol{\zeta}$ 以及潜响应 $\mathbf{y}^*$ 的分布。后验可表示为

$$\Pr(\mathbf{y}^*,\boldsymbol{\zeta},\psi,\boldsymbol{\beta}\mid \mathbf{y},\mathbf{X}) \propto [\Pr(\mathbf{y}\mid\mathbf{y}^*)\Pr(\mathbf{y}^*\mid\boldsymbol{\zeta},\boldsymbol{\beta},\mathbf{X})\Pr(\boldsymbol{\zeta}\mid\psi)]\Pr(\psi)\Pr(\boldsymbol{\beta}) \tag{6.27}$$

在此，归一化常数是观测响应 $\mathbf{y}$ 的边缘分布。方括号内的 $\mathbf{y}$、$\mathbf{y}^*$ 以及 $\boldsymbol{\zeta}$（给定 $\boldsymbol{\beta}$ 和 $\psi$）的联合密度被称为“增强完整数据似然”，其表达式为

$$\Pr(\mathbf{y},\mathbf{y}^*,\boldsymbol{\zeta}\mid\boldsymbol{\beta},\psi,\mathbf{X}) = \prod_{j=1}^{J}\left\{\prod_{i=1}^{n_j}[I(\mathbf{y}_{ij}^*>0)I(\mathbf{y}_{ij}=1)][I(\mathbf{y}_{ij}^*\leqslant 0)I(\mathbf{y}_{ij}=0)]\times \phi(\mathbf{y}_{ij}^*;x_{ij}'\boldsymbol{\beta}+\boldsymbol{\zeta}_j,1)\right\}\phi(\zeta_j;0,\psi)$$

其中，$I(\cdot)$为指示函数。将此表达式代入式(6.27)的后验，很显然，归一化常数不具备闭式，因此很难从后验分布直接模拟。

幸运的是，可使用吉布斯抽样器，因为潜响应、参数以及潜变量的完全条件分布很简单：

(1)$y_{ij}^*$的独立删失正态完全条件，$j=1,\cdots,J,i=1,\cdots n_j$：

$$\Pr(y_{ij}^* \mid \mathbf{X},\mathbf{y},\mathbf{y}^*,\boldsymbol{\zeta},\psi,\boldsymbol{\beta}) = \phi^-(y_{ij}^*;x_{ij}'\boldsymbol{\beta}+\zeta_j,1)^{y_{ij}}\phi^+(y_{ij}^*;x_{ij}'\boldsymbol{\beta}+\zeta_j,1)^{1-y_{ij}}$$

其中，$\phi^-(\cdot)$是一个左删失正态密度，当$y_{ij}^* \leq 0$时等于0；$\phi^+(\cdot)$为右删失正态密度，当$y_{ij}^* > 0$时等于0。为从删失正态来模拟$y_{ij}^*$，我们可从一个正态密度来模拟并忽略落在允许区间以外的抽样。为了避免这种模拟的“浪费”，我们可先产生一个随机均匀(0,1)变量$u_{ij}$。则如果$y_{ij}=1$，我们计算$y_{ij}^* = \Phi^{-1}[1-\Phi(x_{ij}'\boldsymbol{\beta}+\zeta_j)(1-u_{ij})]$；如果$y_{ij}=0$，我们计算$y_{ij}^* = \Phi^{-1}\{[1-\Phi(x_{ij}'\boldsymbol{\beta}+\zeta_j)]u_{ij}\}$。

重要的是，一旦模拟了$y_{ij}^*$，其他随机变量的完全条件则独立于$y_{ij}$了。

(2)具有均值$\bar{\boldsymbol{\mu}}_{\boldsymbol{\beta}} = \bar{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}}[\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}\boldsymbol{\beta}_0 + \sum_{j=1}^{J}\mathbf{X}_j'(\mathbf{y}_j^* - \mathbf{1}_{n_j}\zeta_j)]$，协方差矩阵$\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}} = (\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1} + \mathbf{X}'\mathbf{X})^{-1}$的$\boldsymbol{\beta}$的多元正态完全条件$\Pr(\boldsymbol{\beta} \mid \psi,\boldsymbol{\zeta},\mathbf{y}^*,\mathbf{X},\mathbf{y})$。

(3)$\zeta_j,j=1,\cdots,J$的独立正态完全条件

$$\Pr(\zeta_j \mid \boldsymbol{\beta},\psi,\mathbf{y}^*,\mathbf{X},\mathbf{y}) = \phi[\zeta_j;(\psi+n_j)^{-1}\sum_{i=1}^{n_j}(y_{ij}^* - x_{ij}'\boldsymbol{\beta}),(\psi+n_j)^{-1}]$$

条件均值是由式(7.5)所导出的所谓的经验贝叶斯预测量的一个特例。请注意，如果我们不用$y_{ij}^*$来加强数据，完全条件无法成为多元正态的。

(4)$\psi_i$的逆伽玛完全条件：

$$\Pr(\psi \mid \boldsymbol{\beta},\boldsymbol{\zeta},\mathbf{y}^*,\mathbf{X},\mathbf{y}) = \frac{\left(b+\frac{1}{2}\sum_{j=1}^{J}\zeta_j^2\right)^{a+\frac{J}{2}}}{\Gamma\left(a+\frac{J}{2}\right)}\psi^{-(a+1+\frac{J}{2})} \times \exp\left[-\psi^{-1}\left(b+\frac{1}{2}\sum_{j=1}^{J}\zeta_j^2\right)\right]$$

吉布斯抽样只是简单地通过从上述完全条件分布中抽样，从一些初始数值开始并迭代至一个稳态分布。

请注意，在我们的例子中，吉布斯抽样器的直接应用取决于一个技巧性的设定，包括设定为正态概率和用潜变量来增强数据（如 Tanner and Wong，1987）。在其他情况下必须采用更为复杂的步骤，如对于随机截距对数概率模型，泽戈和卡里姆（Zeger and Karim，1991）建议使用拒斥抽样，斯皮格哈特等（Spiegelhalter et al.，1996）使用自适应拒斥抽样（Gilks and Wild，1992），而布朗尼和德雷珀（Browne and Draper，2004）则使用混合 Metropolis-Hastings 办法。

**交替插值后验算法**

对于包含交叉随机效应的广义线性模型，克雷顿和拉斯巴什于1999年提出了一种特殊类型的 MCMC 方法（Clayton and Rasbash，1999）。他们的算法是基于唐纳和黄（Tanner and Wong，1987）的插值后验（IP）的算法，此算法在插值（I）步进和后验（P）步进之间迭代。该算法类似吉布斯抽样，除了在 P 步进上，整个参数向量是在给定潜变量条件下从其后验条件分布中取样，而不是单分量。

该算法可以概述如下：

I 步进:给定 $\mathbf{y}$,$\boldsymbol{\vartheta}^{r-1}$以及 $\mathbf{X}$(数据增强),从 $\boldsymbol{\zeta}$ 的后验分布中抽取一个样本 $\boldsymbol{\zeta}^r$。

P 步进:给定 $\boldsymbol{\zeta}^r$,$\mathbf{y}$ 和 $\mathbf{X}$,从 $\boldsymbol{\vartheta}$ 的后验分布中抽取一个样本 $\boldsymbol{\vartheta}^r$。

正如吉布斯抽样,该算法运行直到达到稳态分布(老化期),并且参数通过它们的均值得以估计,即

$$\mathrm{E}(\boldsymbol{\vartheta} \mid \mathbf{y},\mathbf{X}) \approx \bar{\boldsymbol{\vartheta}} \equiv \frac{1}{R}\sum_{r=1}^{R}\boldsymbol{\vartheta}^r$$

在 I 步进,使用通常的固定参数的潜变量的经验贝叶斯后验分布(见 7.3 节),但此处参数集合被设为等于 $\boldsymbol{\vartheta}^r$ 而非最大似然估计。在 P 步进,在前面迭代中所抽取的随机效应被视为固定偏移量。然后用多元正态分布对参数的后验分布近似,此多元正态分布的均值由最大似然估计 $\hat{\boldsymbol{\vartheta}}^r$ 给予(将 $\boldsymbol{\zeta}^r$ 作为偏移量),其协方差矩阵 $\hat{\boldsymbol{\Sigma}}^r$ 由海森矩阵导出。如果假定所有参数 $\boldsymbol{\vartheta}$ 都是均匀先验,那么该近似"抽样分布"将逼近真实贝叶斯后验。参数估计的方差(用劳-布莱克威尔法(Rao-Blackwellization))估计为

$$\mathrm{Var}(\boldsymbol{\vartheta} \mid \mathbf{y},\mathbf{X}) \approx \frac{1}{R}\sum_{r=1}^{R}\hat{\boldsymbol{\Sigma}}^r + \frac{1}{R}\sum_{r=1}^{R}(\boldsymbol{\vartheta}^r - \bar{\boldsymbol{\vartheta}})(\boldsymbol{\vartheta}^r - \bar{\boldsymbol{\vartheta}})'$$

也就是插值内部及相互之间方差的总和。克雷顿和拉斯巴什(Clayton and Rasbash,1999)指出,当**单个**参数是从吉布斯抽样的条件中抽取的时候(而不是这里的整个参数向量 $\boldsymbol{\vartheta}$),这个劳-布莱克威尔法不能用于吉布斯抽样,因为条件协方差($\hat{\boldsymbol{\Sigma}}^r$ 的非对角线元素)在这种情况下是不存在的。

克雷顿和拉斯巴什(Clayton and Rasbash,1999)认为,理想的 P 步进应由两部分组成:①使用 REML 以获取方差和协方差参数的近似后验,并从后验中抽取;②固定参数的后验通过多元正态分布来逼近,其中,多元正态分布的均值和方差源于 ML 解,设定方差参数等同于①中的抽取。

克雷顿和拉斯巴什(Clayton and Rasbash,1999)利用该算法来估计包含交叉随机效应的模型。在他们的应用中,妇女进行多次来自不同捐献者的人工授精,每一个尝试的响应即为成功或失败。由于每个捐献者的精子也授予不同妇女,因此,妇女-特定的随机效应与捐献者-特定的随机效应互相交叉。在他们的交替插值后验(AIP)算法中,克雷顿和拉斯巴什交替使用捐献者和妇女部分的数据。在捐献者部分,进行一次 IP 迭代,将妇女-特定的效应视为偏移量,并抽取一个新的捐献者-特定的随机效应样本。在妇女部分,捐献者特定的随机效应在一次 IP 迭代中被视为偏移量,并更新妇女-特定的随机效应。各个部分交替一直到收敛。

通过从两个部分中获得单独的对模型参数的估计,则可对算法的收敛性进行评估。最后的估计则是对两个部分的均值。埃克沙德和克雷顿(Ecochard and Clayton,2001)建议并行地运行两个部分。在 11.4 节,我们使用一种修改的 AIP 算法,用以对空间相关的随机效应进行疾病制图,该随机效应模型无法使用积分方法。

### 6.11.6 MCMC 方法的利弊

MCMC 方法的方法已被用于各种各样的潜变量模型,包括广义线性混合模型(如 Zeger and Karim,1991;Clayton,1996)、多层次模型(如 Browne,1998)、协变量测量误差模型(如 Richardson and Gilks,1993)、疾病制图(如 Mollie,1996)、多层次因子模型(如 Goldstein and Browne,2002)和多层次题器响应模型(如 Ansari and Jedidi,2000;Fox and Glas,2001)。许多应用可见于康登(Congdon,2001)和斯皮格哈特等(Spiegelhalter et al.,1996bc)。我们使用 MCMC 方法,在 9.5 节进行了元分析,在 11.4 节中做了疾病制图。

MCMC 方法的一个重要的优点是，它可用于估计复杂的模型，而其他方法要么不适用，要么工作不佳。另一个优点是，后验分布的任何特征都可以在稳态模拟值的基础上进行分析，如在制度比较中，随机效应的后验均值和次序分位点（如 Goldstein and Spiegelhalter，1996）。

在使用 MCMC 方法中，有一些或多或少有争议的问题必须解决。首先是**老化期**，由于对初始值的依赖，必须确定需要抛弃的初始迭代的次数。不幸的是，监测一个分布的收敛比一个点的收敛要困难得多（Gelman and Rubin，1996）。一个流行的办法是使用一个任意大数值，或运行若干个具有不同初始状态的链，以评估收敛。然而，在关于需要依赖的**链的数量**上面，则存在众多争议，如使用若干长链、一个极长链和若干短链（后者似乎被误导了）。

另一个问题是决定何时**停止**链，以确保可接受的估计精度。当存在缓慢的混合时，即当链“通过目标分布的瓶颈”运行缓慢时，尤其难以判断估计的收敛（Gelman and Rubin，1996）。当混合很差时，链需要运行很长的时间以获取准确估计。

虽然贝叶斯方法可用于辨识，但采用 MCMC 方法的实现将难以发现辨识不足的问题（如 Keane，1992）。其中一个原因是，与使用梯度方法的最大化相比，平缓后验将是估计的一个自然副产品，因此，它难以被发现。链的混合不足将导致错误地认为一个不足以辨识的参数被精确地估计了。

另一个问题涉及在随机效应模型中，方差参数的非信息先验的设定（如 Natarajan and Kass，2000；Hobert，2000）。例如，霍伯特（Hobert，2000）指出，泽戈和卡里姆（Zeger and Karim，1991）在吉布斯抽样中所用的先验可能会导致一个不正确的后验分布（不是积分为 1），尽管所有的完全条件都是正确的。不幸的是，从吉布斯的输出中，这个问题可能并不明显。此外，即使得到了正确的后验，但在任何正式的意义上，先验也不是非信息的，这意味着先验在实际上驱动推断。这个问题也存在于“扩散的”正确先验中，这些“扩散的”正确先验通常是正确的共轭先验而又几乎是不正确的。此外，这些先验可能导致吉布斯抽样器收敛速度很慢。霍伯特（Hobert，2000）得出结论认为，在没有先验信息的情况下，为方差参数选择一个先验是目前贝叶斯学派的真正难题。霍伯特同时也提出需关注对估计的标准误的理论属性。

## 6.12 总 结

通过显性积分的完全最大似然的优点是：①在数据为随机性缺失（MAR）时，估计具有一致性；②对基于似然的推断，似然可用。精度可以通过增加积分点或蒙特卡洛重复得到改善和评估。此外，该方法适用于一般模型框架。其缺点是计算效率低下。随潜变量的增加，高斯-埃尔米特方法对计算的需求尤其高。劳登布什等（Raudenbush et al.，2000）所提出的六阶拉普拉斯近似似乎是非常有效的，并在很多情况下可能是足够准确的。

对于一般的多元正态的潜响应模型，穆森的有限信息方法也是一个非常好的选择。然而，群组规模必须是（或接近是）常数，且仅有少量缺失数据或少量缺失数据模式。此估计方法在计算上非常有效，并能给出非常接近最大似然的估计，但小样本复杂模型除外。令人惊讶的是，在多层次建模文献里，这个方法很少受到关注。一个相关的有限信息方法是“伪似然”方法，见 le Cessie and van Houwelingen（1994）以及 Geys et al.（2002）。

MQL 和 PQL 也是计算效率非常高的，无论潜变量的数量是多少。不幸的是，这些方

法有时会产生严重偏误的估计。不同于基于积分的方法,该方法的精度不能逐步提高,因此难以评估它们的可靠性。这可以用库克所提出的偏差矫正的参数自助法来加以纠正(另见 Goldstein,2003)。MQL 和 PQL 目前局限于广义线性混合模型。

MCMC 方法允许对类型非常广泛的模型进行估计,并已越来越受欢迎。然而,这种灵活性可能导致设定过于复杂的模型,这些模型有可能无法辨识,并很难对先验分布的影响进行评估。这些问题往往由于对模型的不充分描述、估计列表的缺乏及其标准误而被加剧。此外,收敛可能是缓慢的并很难监测。

对于潜变量模型被认为不太适当的应用,采用 GEE 或固定效应方法将可能是有用的。

但是,文献中缺少对不同的估算方法的性能表现进行比较的模拟研究,如斯科隆多(Skrondal,2000)所提出的,缺少系统地评估不同因素对算法表现的影响,如群组规模、组内相关、样本规模等。Breslow(2003)、McCulloch and Searle(2001)、Rabe-Hesketh et al.(2002)以及 Goldstein(2003)对潜变量模型的不同估计方法进行了很好的综述。

附录:部分软件及参考

对于本章所讨论的每个主要方法,我们列出部分参考以及其软件实现。我们不提供这些软件的地址或链接,因为这些信息更新得很快,也很容易从互联网获得。

• 闭式边缘似然(6.2 节):

——参考:Browne and Arminger(1995)潜变量结构方程建模(也讨论了其他方法)。

——软件:

* AMOS(Arbuckle and Wothke,1999,2003)。

* EQS(Bentler,1995)。

* LISREL(Jöreskog and Sörbom,1994;Jöreskog et al.,2001)。

* MECOSA(Arminger et al.,1996)。

* MX(Neale et al.,2002)。

• 拉普拉斯近似:

——参考:Tierney and Kadane(1986)、Tanner(1996,第 3.2 节)以及 Raudenbush et al.(2000)。

——软件:HLM,多层次广义线性混合模型中的六阶拉普拉斯(Raudenbush et al.,2001)。

• 高斯-埃尔米特积分:

——参考:Stroud and Secrest(1966)以及 Davis and Rabinowitz(1984)。

——软件:

* aML,多层次及多过程模型(Lillard and Panis,2000)。

* BILOG-MG,二分对数概率题器响应模型(如 Du Toit,2003)。

* EGRET,双层次随机截距模型(EGRET Windows 用户手册,2000)。

* gllamm,Stata 中的广义线性潜变量及混合模型(Rabe-Hesketh et al.,2001b,2004c)。

* LIMDEP,双层次随机截距模型(Greene,2002a)。

* MIXNO,双层次多项式对数概率模型(Hedeker,1999)。

* MIXOR,双层次定序对数概率及正态概率回归(Hedeker and Gibbons,1996a)。

* MIXREG,包含自相关误差的双层次线性混合模型(Hedeker and Gibbons,1996b)。

* MIXSUR 和 MIXPREG,计数及离散时间存续模型(由 Hedeker)。

* MULTILOG，多项式对数概率题器响应模型（如 Du Toit,2003）。

* PARSCALE，定序正态概率和对数概率题器响应模型（如 Du Toit,2003）。

* SABRE，双层次广义线性混合模型（Francis et al.,1996）。

* Stata 的 xt 命令，双层次随机截距模型（Stata Cross-Sectional Time-Series,2003；StataCorp,2003）。

* TESTFACT，探索性多维正态概率因子模型(Bock et al.,1999;Du Toit,2003)。

• 自适应积分：

——参考：Pinheiro and Bates（1995）、Bock and Schilling（1997）、Evans and Swartz（2000）以及 Rabe-Hesketh et al.（2002,2004b）。

——软件：

* gllamm，Stata 中的广义线性潜变量及混合模型（Rabe-Hesketh et al.,2001b,2002,2004c）。

* NLMIXED,SAS 中的双层次广义线性混合模型（Wolfinger,1999）。

* TESTFACT,探索性多维正态概率因子模型（Bock et al.,1999;Du Toit,2003）。

• 蒙特卡洛积分：

——参考：Train（2003）、Cappellari and Jenkins（2003）以及 Gourieroux and Montfort（1996）.

——软件：

* mvprobit,Stata 中的多元正态概率回归(Cappellari and Jenkins,2003)。

* NLOGIT,包含随机效应的多项式对数概率模型、正态概率模型以及其他离散选择模型（Greene,2002b）。

* Ox 中的 DCM 命令（Eklöf and Weeks,2003）。

* GAUSS 中的面板数据混合对数概率估计命令（Train et al.,1999）。

• EM 算法：

——参考：Tanner(1996)、Schafer(1997)、McLachlan and Krishnan(1997)以及 Little and Rubin(2002)。

——软件（部分例子）：

* HLM,多层次广义线性混合模型（Raudenbush et al.,2001）。

* Latent GOLD,潜类模型及相关的不同响应类型的模型(Vermunt and Magidson,2000,2003a)。

• 梯度法：

——参考：Judge et al.（1985）、Fletcher(1987)、Gould et al.（2003）、Thisted(1987)以及 Everitt(1987)。

——用于最大化任意似然的软件：

* Stata 的 ml 命令(Gould et al.,2003)。

* GAUSS 的扩展程序 CML(Schoenberg,1996)。

• 有限信息：

——参考：Skrondal(1996)、Küsters(1987)以及 Muthén and Satorra(1995)。

——软件：

* Mplus(Muthén and Muthén,1998,2003)。

* MECOSA(Arminger et al.,1996)。

•（边缘和罚）伪似然：

——参考:Goldstein(2003)以及 Breslow(2003)。

——软件:

* HLM,多层次广义线性混合模型 (Raudenbush et al.,2001)。

* MLwiN,广义线性混合模型 (Rasbash et al.,2000)。

* GLIMMIX,SAS 中的广义线性混合模型 (SAS/Stat 用户手册,第 8 版,2000)。

* GLMM,Genstat 中的广义线性混合模型 (Payne,2002)。

* glmmPQL,S 和 R 中的广义线性混合模型 (Venables and Ripley,2002)。

• 分级似然 (h-似然):

——参考:Lee and Nelder(1996,2001)。

——软件:

* Genstat 中的 HG 过程 (Payne,2002)。

• 广义估计方程:

——参考:Pickles(1998)、Hardin and Hilbe(2002)以及 Molenberghs(2002)。

——软件:

* Stata 的 xtgee 命令(Stata Cross-Sectional Time-Series,2003)。

* SAS 的 GENMOD(SAS/Stat 用户手册,第 8 版,2000)。

* Genstat 的 GEE in(Payne,2002)。

• 联合最大似然:

——参考:Hambleton and Swaminathan(1985) 关于 IRT 模型;Lancaster(2000) 关于伴随参数问题。

• 条件最大似然:

——参考:Clayton and Hills(1993)、Breslow and Day(1980)、Hamerle and Rönning(1995)以及 Cameron and Trivedi(1998)。

——软件:

* EGRET(EGRET Windows 用户手册,2000)。

* Stata 的 clogit,xtpoisson 等命令 (Stata Cross-Sectional Time-Series,2003)。

* SAS 的 LOGISTIC(SAS/Stat 用户手册,第 8 版,2000)。

• Bayes:

——参考:Gelman et al. (2003) 关于"实用"贝叶斯统计;Casella and George(1992)关于吉布斯抽样器;Chib and Greenberg(1995)关于 Metropolis-Hastings 算法;Rubin(1991)、Gelman and Rubin(1996)、Gilks et al. (1996)以及 Gilks(1998)关于不同的 MCMC 方法;Zeger and Karim(1991)、Albert(1992)、Albert and Chib(1993)、Dellaportas and Smith(1993)、Clayton(1996b)、Arminger and Muthén(1998)以及 Fox and Glas(2001)关于广义线性混合模型、结构方程模型以及 IRT 模型的 MCMC 方法。

——软件:

* BUGS 和 WinBUGS,一般贝叶斯以及多层贝叶斯模型(Spiegelhalter et al.,1996abc;另见 Congdon,2001)。

* GLMM 吉布斯,R 中的吉布斯抽样广义线性混合模型 (Myles and Clayton,2001)。

* MLwiN,广义线性混合模型以及多层次因子模型(Rasbash et al.,2000;另见 Goldstein and Browne,2002)。

* Gauss 中的无边界分布混合对数概率程序(Train,2002)。

# 7 潜变量赋值

## 7.1 介　绍

在本章,我们讨论对于单个群组的潜变量进行赋值的方法。群组是指纵贯性测量模型中的对象,或者多层次模型中的学校等。对于连续潜变量,我们将赋值称为**潜计分**(latent scoring)(因子计分或随机效应计分),对于离散潜变量则称为**分类**(classification)。请注意,这个术语并未区分不同的响应类型;例如因子计分就包括带二项响应的 IRT 模型的计分。

在一些研究中,计分与分类是潜变量模型的主要目标,典型的例子如 IRT 模型的能力计分以及基于潜类模型的医疗诊断。其他例子还包括疾病制图,小域估计以及制度效用评估。

在前面一章中,我们分析了基本参数 $\boldsymbol{\vartheta}$ 的估计问题。在此我们假设这些参数已经被估计为 $\hat{\boldsymbol{\vartheta}}$,所得到的结构参数估计就为 $\hat{\boldsymbol{\vartheta}}=\mathbf{h}(\hat{\boldsymbol{\vartheta}})$。有些情况下,需要根据一个很大的**校准样本**(calibration sample)的数据来估计这些参数,而这个样本不包含需要计分的这个群组,如在教育测验中。这个方法的优点在于参数估计非常精确。然而,将估计结果用于不同的总体将会出现问题。

与上一章不同,我们将完全关注于频率学派方法,将所估计的结构参数当做已知。从贝叶斯方法的角度来看,参数以及潜变量都会被当做随机变量,因此,在参数估计和潜计分或分类之间没有根本差别。从经验贝叶斯的角度来看,关于潜变量的推论基于给定参数(代入估计值)下的条件后验分布。在 7.2 节将讨论经验贝叶斯后验分布。

我们将考虑 3 种对潜变量赋值的方法,这些方法都来自于一般统计学原理。在 7.3 节讨论使用经验贝叶斯(EB)(也被称为期望后验(expected a posteriori, EAP))的**预测法**,在 7.4 节讨论使用经验贝叶斯众数(EBM)(也被称为众数后验(modal a posteriori, MAP))的预测法。使用最大似然法(ML)的**估计法**在 7.5 节中介绍。对于潜计分来说,经验贝叶斯是最常用的方法,而经验贝叶斯众数则是分类中最为常用的方法。

对于每种计分方法,我们将对比不同类型的潜在得分。这对于经验贝叶斯预测来说特别有用,因为在其文献中对不同类型的方差的意义存在相当程度的混淆。

除了对一般模型讨论计分和分类以外,我们还研究特殊的模型,如不包含结构部分(如因子模型和随机效应模型),但包含多元正态潜变量和响应的模型,为了简明起见,我们统称为“线性案例”。许多被人所熟悉的计分模型都可看做是这种一般方法的特殊情

形,这使得我们更加有动力去理解这些方法。对“线性案例”的**闭式表达式**也有助于讨论一些概念,如“**收缩**(shrinkage)”。为了帮助理解,将介绍一个双层次随机截距模型例子。对于更加复杂的线性模型,为了“看见”在实际例子中发生了什么,我们发现将数值代入公式更有助于学习。在7.6节中,我们证明这3种计分方法在“线性案例”中的相关关系。

由于对于假设构念的潜计分,通常使用特定的方法,在7.7节简单讨论了这些方法。在7.8节,我们探索潜计分和分类的更加广泛的应用。对潜变量赋值的不同方法的软件实现,我们将它们放在附录中介绍。

如果没有特殊说明,我们的讨论限定于连续潜变量。另外,我们也将我们的显式处理限定于对干扰项或残差$\boldsymbol{\zeta}$(均值为0),而非对$\boldsymbol{\eta}$,而$\boldsymbol{\eta}$除了包含干扰项外,还可能包含对观测协变量及潜变量的回归。这些处理只是为了简化标记,以及因为干扰项本身就非常值得关注。由于结构方程为线性的,对应的$\boldsymbol{\eta}$的得分可通过将干扰项的得分代入潜变量的简化式(4.18)而得到。如果不存在结构模型,那么就可直接得到$\boldsymbol{\zeta}=\boldsymbol{\eta}$。

## 7.2 后验分布

当需要对潜变量赋值时,频率学派通常求诸于贝叶斯理论。其原因在于潜变量可被理解为带有“先验”分布$h(\boldsymbol{\zeta};\boldsymbol{\theta})$的随机“参数”(如随机效应),使得模型看起来类似于贝叶斯模型。一个很重要的区别在于6.11节所讨论的完整贝叶斯法除了将干扰项$\boldsymbol{\zeta}$假设为先验以外,还对结构参数$\boldsymbol{\theta}$也做了先验假设。在这种情形下,**$\boldsymbol{\zeta}$先验的参数的先验**,如潜变量的方差和协方差,则被认为是**超先验**。贝叶斯法和潜计分一样,其参数估计基于给定响应$\mathbf{y}$条件下的后验分布。与此相关的计分后验就为关于$\boldsymbol{\theta}$的边缘,而参数估计的后验就为关于$\boldsymbol{\zeta}$的边缘。

至于潜计分,频率学派通常采用经验贝叶斯法。它们用最大似然法(或其他方法)估计参数$\boldsymbol{\theta}$,但是在预测上则依赖于在给定估计参数$\boldsymbol{\zeta}$条件下的潜变量的条件后验分布。为了方便起见,我们都将使用贝叶斯术语;当$\boldsymbol{\zeta}$不是所估计的模型参数而是固定的常数时,这在技术上是正确的。

关于干扰项$\boldsymbol{\zeta}$,有3个不同的信息来源。其第一个信息为先验分布$h(\boldsymbol{\zeta};\hat{\boldsymbol{\theta}})$,表示我们对潜变量的先验知识,当它们为连续变量时,通常设定为多元正态。第二个信息由“测量”潜变量的观测而得到的响应$\mathbf{y}$提供。第三个信息,尽管并不总是存在,为响应或结构模型中的协变量$\mathbf{X}$。

请注意,对于潜计分和分类,是否必须使用模型中的协变量信息并没有严格约定。例如大多数人认为在心脏病诊断中应该使用年龄、性别以及种族这样的协变量,如果这能够降低错误诊断的风险的话。相反,在教育测试中使用这些协变量信息则很可能被认为是“政治上不正确”或不公平,尽管有助于提高能力评估的质量(另见9.4节)。

将这些信息来源合并使用的一个自然方式是通过$\zeta$的**后验分布**$\omega(\boldsymbol{\zeta}|\mathbf{y},\mathbf{X};\hat{\boldsymbol{\theta}})$,$\boldsymbol{\zeta}$的分布由$\mathbf{y}$和$\mathbf{X}$的数据更新或给定。因此,该后验就提供了关于潜得分推论的自然前提(另见 Bartholomew, 1981)。根据贝叶斯定理,可得

$$\omega(\boldsymbol{\zeta}\mid \mathbf{y},\mathbf{X};\hat{\boldsymbol{\theta}}) = \frac{\Pr(\mathbf{y},\boldsymbol{\zeta}\mid \mathbf{X};\hat{\boldsymbol{\theta}})}{\Pr(\mathbf{y}\mid \mathbf{X};\hat{\boldsymbol{\theta}})}$$

在通常的$L$-层次模型中,在顶层群组$z$中的每一个第一层次单元的线性估计量依赖

于该群组中潜变量向量$\boldsymbol{\zeta}_{z(L)}$的几个元素。如,在三层次随机截距模型中,正如3.2节的图3.2所示,每个响应同时依赖于第二和第三层次的随机截距。由此说明,给定响应提供了多于一个潜变量的信息,因此,在后验分布下,潜变量通常是因变量。因而,分析给定响应$\mathbf{y}_{z(L)}$以及顶层群组$z$的全部协变量$\mathbf{X}_{z(L)}$下的$\boldsymbol{\zeta}_{z(L)}$的联合后验分布,是非常有帮助的,即

$$\omega(\boldsymbol{\zeta}_{z(L)} \mid \mathbf{y}_{z(L)}, \mathbf{X}_{z(L)}; \hat{\boldsymbol{\theta}}) = \frac{\Pr(\mathbf{y}_{z(L)}, \boldsymbol{\zeta}_{z(L)} \mid \mathbf{X}_{z(L)}; \hat{\boldsymbol{\theta}})}{\Pr(\mathbf{y}_{z(L)} \mid \mathbf{X}_{z(L)}; \hat{\boldsymbol{\theta}})}$$

为了加深印象,请考虑一个简单的双层次随机截距模型。对于给定群组$j$,上面公式分子中的随机截距$\zeta_j(2)$的联合密度以及响应$\mathbf{y}_{z(L)}$可写为

$$h(\zeta_{j(2)}; \hat{\boldsymbol{\theta}}) \prod g^{(1)}(y_{ij} \mid \mathbf{x}_{ij}, \zeta_{j(2)}; \hat{\boldsymbol{\theta}})$$

在此,我们利用了给定$\zeta_{j(2)}$,$y_{ij}$条件独立。另外,在分母中的响应的边缘密度只是该联合密度关于随机截距的积分,因此该后验密度就成为

$$\omega(\zeta_{j(2)} \mid \mathbf{y}_{j(2)}, \mathbf{X}_{j(2)}; \hat{\boldsymbol{\theta}}) = \frac{h(\zeta_{j(2)}; \hat{\boldsymbol{\theta}}) \prod g^{(1)}(y_{ij} \mid \mathbf{x}_{ij}, \zeta_{j(2)}; \hat{\boldsymbol{\theta}})}{\int h(\zeta_{j(2)}; \hat{\boldsymbol{\theta}}) \prod g^{(1)}(y_{ij} \mid \mathbf{x}_{ij}, \zeta_{j(2)}; \hat{\boldsymbol{\theta}}) \mathrm{d}\zeta_{j(2)}}$$

由此可见,该后验无法用闭式来表达,因为在分母上的积分并没有解析解。由于分母只是第$j$个第二层次单元的似然贡献,因此,该问题与6.3节所讨论的对边缘似然的估计一样。

在4.2.2节已介绍过用$\boldsymbol{\eta}_{(L)}$和$\boldsymbol{\zeta}_{(L)}$将通常的$L$-层次模型写为双层次模型,上述表达式也可用于$L$-层次模型,即

$$\omega(\boldsymbol{\zeta}_{z(L)} \mid \boldsymbol{y}_{z(L)}, \boldsymbol{X}_{z(L)}; \hat{\boldsymbol{\theta}}) = \frac{h(\boldsymbol{\zeta}_{z(L)}; \hat{\theta}) \prod g^{(1)}(y_{ij\cdots z} \mid \boldsymbol{x}_{ij\cdots z}, \boldsymbol{\zeta}_{z(L)}; \hat{\boldsymbol{\theta}})}{\int h(\boldsymbol{\zeta}_{z(L)}; \hat{\theta}) \prod g^{(1)}(y_{ij\cdots z} \mid \boldsymbol{x}_{ij\cdots z}, \boldsymbol{\zeta}_{z(L)}; \hat{\boldsymbol{\theta}}) \mathrm{d}\boldsymbol{\zeta}_{z(L)}} \tag{7.1}$$

其中,乘积为在第$z$个$L$层次单元内部的所有第一层次的单元。然而,分母的积分维度非常高,数值积分必须利用条件依赖结构,用6.3节的式(6.2)来递归估计积分。

在该"线性案例"中,根据条件多元正态密度(如 Anderson,2003)的标准结果可得,后验密度为**多元正态**(对于期望向量和协方差矩阵请见式(7.3)和式(7.7))。对于其他响应类型,根据**贝叶斯中心极限定理**(如 Carlin and Louis,1998),随着群组中的单元数量增加,后验密度趋向于多元正态。

最后,再考虑一个**离散**随机截距模型,对于原点$\zeta^{(2)} = e_c, c = 1, \cdots, C$。其后验概率为$\pi_{cj}$。其截距等于$e_c$的后验概率为

$$\omega(e_c \mid \mathbf{y}_{j(2)}, \mathbf{X}_{j(2)}; \hat{\boldsymbol{\theta}}) = \frac{\pi_{cj} \prod g^{(1)}(y_{ij} \mid \mathbf{x}_{ij}, e_c; \hat{\boldsymbol{\theta}})}{\sum_{k=1}^{C} \pi_{kj} \prod g^{(1)}(y_{ij} \mid \mathbf{x}_{ij}, e_k; \hat{\boldsymbol{\theta}})}$$

在依据单一检验结果$y_j$的医学诊断中,如果$y_j = 1$(正的测试结果)且$c = 2$(疾病存在),则这些概率被称为"阳性预测值";如果$y_j = 0$(负的测试结果)且$c = 1$(疾病不存在),则被称为"阴性预测值"。在这种情形中,先验概率为疾病的流行程度,即表示医生在看到检验结果之前对患者诊断的知识。如果检验有用,后验概率将比先验概率更加接近于0或1。请见9.3节心肌梗塞(心脏病发作)诊断的例子。

## 7.3 经验贝叶斯(EB)

### 7.3.1 经验贝叶斯预测

毫无疑问,对于因子和随机效应计分,经验贝叶斯预测是最为广泛使用的方法。潜变量$\boldsymbol{\zeta}_{(L)}$的经验贝叶斯预测(见 Efron and Morris,1973,1975;Morris,1983)为它们代入参数估计$\hat{\boldsymbol{\theta}}$后的后验均值,即

$$\tilde{\boldsymbol{\zeta}}_{z(L)}^{\mathrm{EB}} = \mathrm{E}(\boldsymbol{\zeta}_{z(L)} \mid \mathbf{y}_{z(L)}, \mathbf{X}_{z(L)}; \hat{\boldsymbol{\theta}})$$

当先验分布为参数分布时,该预测被标记为"参数经验贝叶斯"。

之所以用"经验贝叶斯"这样一个概念,是因为,正如前面所提到的,贝叶斯定理是通过代入所估计的模型参数而被频率学派所采纳。贝叶斯学派则会去获得潜变量的后验分布,即关于$\boldsymbol{\theta}$的边缘,而非简单地将$\boldsymbol{\theta}$替代为估计值。很显然,贝叶斯方法和经验贝叶斯方法基于不同的哲学基础,由此林莱(Lindley,1969)写到:

*没有什么比经验贝叶斯更加不贝叶斯得了。*

然而,在贝叶斯和经验贝叶斯这两个术语之间仍然存在相当程度的混淆。这一点从计量心理学文献中(如 Bock and Aitkin,1981;Bock and Mislevy,1982;Bock,1983,1985;Muraki and Engelhard Jr. ,1985),"期望后验"预测量(EAP)这个名词的流行也反映出来,这个名词似乎意味着它是真正的贝叶斯估计。

尽管贝叶斯推论和经验贝叶斯推论在理论上不同,但在大样本中且(或者)(超)先验不清楚的情况下,当$\hat{\boldsymbol{\theta}}$一致,如果似然决定了$\boldsymbol{\theta}$的(超)先验,对参数估计进行替代的效应非常微小。但是,在小样本或中等规模样本的情形中,其后果仍不清楚。迪莱和林莱(Deely and Lindley,1981)提供了一个可靠的理论结果,他们指出经验贝叶斯预测量为贝叶斯预测量的一阶近似。

经验贝叶斯预测量可通过**加总二次损失函数**(summed quadratic loss function)来证明,该函数被定义为预测量$\tilde{\zeta}_{(L)}$的未加权的平方差之和。去掉下标 z,得到

$$\mathrm{L}^{\mathrm{EB}}(\boldsymbol{\zeta}_{(L)}, \tilde{\boldsymbol{\zeta}}_{(L)}) = (\boldsymbol{\zeta}_{(L)} - \tilde{\boldsymbol{\zeta}}_{(L)})'(\boldsymbol{\zeta}_{(L)} - \tilde{\boldsymbol{\zeta}}_{(L)})$$

将这些参数当做已知,经验贝叶斯预测量使期望后验损失最小化:

$$\int (\boldsymbol{\zeta}_{(L)} - \tilde{\boldsymbol{\zeta}}_{(L)})'(\boldsymbol{\zeta}_{(L)} - \tilde{\boldsymbol{\zeta}}_{(L)})\omega(\boldsymbol{\zeta}_{(L)} \mid \mathbf{y}_{(L)}, \mathbf{X}_{(L)}; \hat{\boldsymbol{\theta}})\mathrm{d}\boldsymbol{\zeta}_{(L)} \tag{7.2}$$

请注意,式(7.2)也可采用直观的解释,即与**预测的后验均方差**成比例,其中期望值取遍后验分布。

希尔勒等(Searle et al.,1992:262)证明,如果把**参数被当做已知**,经验贝叶斯预测量也使得预测均方差在 y 的样本分布上最小化。对于频率学派来讲,这个结果可能比用后验损失方法来证明的(经验)贝叶斯更为有用。麦克拉奇和希尔勒(McCulloch and Searle,2001:257-258)强调指出,将所估计的参数代入经验贝叶斯预测量是一种纯粹实用主义的做法,没有任何统计学基本原理的道理。

很显然,对于真实的离散潜变量来说,其预测值必须限于潜类别的范围,因此,均方差损失函数毫无意义。因此,经验贝叶斯预测不应用在这种情形中。但是,在非参数最大似然估计(NPMLE)中,离散分布被理解为可能的连续潜变量分布的非参数估计量,因

此,经验贝叶斯预测可用在非参数最大似然估计中。基于 NPMLE 的经验贝叶斯预测被 Clayton and Kaldor(1987)、Laird(1982)和 Rabe-Hesketh et al.(2003a)以及其他一些学者所采用。后者的论文发现,对于偏斜的潜变量分布来说,基于 NPMLE 的 EB 预测值优于基于正态的预测值。请见 11.4 节的基于 NPMLE 的经验贝叶斯在疾病制图中的应用例子。

对于包含随机截距 $\eta_j^{(2)} = \zeta_j^{(2)}$ 的双层次随机截距模型:

$$\nu_{ij} = \mathbf{x}_{ij}'\boldsymbol{\beta} + \zeta_j^{(2)}$$

经验贝叶斯预测量则为

$$\tilde{\zeta}_{j(2)}^{\mathrm{EB}} = \frac{\int \zeta_j^{(2)} h(\zeta_j^{(2)};\hat{\boldsymbol{\theta}}) \prod_i g^{(1)}(y_{ij} \mid \mathbf{x}_{ij}, \zeta_j^{(2)};\hat{\boldsymbol{\theta}})\,\mathrm{d}\zeta_j^{(2)}}{\int h(\zeta_j^{(2)};\hat{\boldsymbol{\theta}}) \prod_i g^{(1)}(y_{ij} \mid \mathbf{x}_{ij}, \zeta_j^{(2)};\hat{\boldsymbol{\theta}})\,\mathrm{d}\zeta_j^{(2)}}$$

在通常的 $L$ 层次模型中,在层次 $l$ 上的潜变量 $\zeta^{(l)}$ 的经验贝叶斯预测可用递归法得

$$\mathrm{E}(\zeta^{(l)} \mid \boldsymbol{\zeta}^{([l+1]+)}) = \frac{\int \zeta^{(l)} h(\zeta^{(l)}) \prod g^{(l-1)}\,\mathrm{d}\zeta^{(l)}}{\int h(\zeta^{(l)}) \prod g^{(l-1)}\,\mathrm{d}\zeta^{(l)}}$$

以及

$$\mathrm{E}(\zeta^{(l)} \mid \boldsymbol{\zeta}^{([l+k+1]+)}) = \frac{\int \mathrm{E}(\zeta^{(l)} \mid \boldsymbol{\zeta}^{([l+k]+)}) h(\boldsymbol{\zeta}^{(l+k)}) \prod g^{(l+k-1)}\,\mathrm{d}\boldsymbol{\zeta}^{(l+k)}}{\int h(\boldsymbol{\zeta}^{(l+k)}) \prod g^{(l+k-1)}\,\mathrm{d}\boldsymbol{\zeta}^{(l+k)}}$$

其中,$k=1,\cdots,L-l$,并且我们将 $g^{(l)}(\mathbf{y}_{(l)} \mid \mathbf{X}_{(l)}, \boldsymbol{\zeta}^{([l+1]+)};\hat{\boldsymbol{\theta}})$ 写为 $g^{(l)}$ 以简化标记。一般来说,用解析积分可得到经验贝叶斯预测。在 6.3 节所讨论的各种数值积分法均可使用,如 6.3.2 节的自适应积分法。对于离散分布,可直接将积分替换为加总。

"线性案例":我们现在来考虑没有结构模型的"线性案例",即 $\boldsymbol{\eta}_{z(L)} = \boldsymbol{\zeta}_{z(L)}$,按照式(4.3),设定如下:

$$\mathbf{y}_{z(L)} = \mathbf{X}_{z(L)}\boldsymbol{\beta} + \boldsymbol{\Lambda}_{z(L)}\boldsymbol{\zeta}_{z(L)} + \boldsymbol{\epsilon}_{z(L)}$$

在此,潜变量 $\boldsymbol{\zeta}_{z(L)}$ 为多元正态分布,其估计的协方差矩阵为 $\hat{\boldsymbol{\Psi}}_{(L)}$;干扰项 $\boldsymbol{\epsilon}_{z(L)}$ 为多元正态分布,其估计的对角线协方差矩阵为 $\hat{\boldsymbol{\Theta}}_{z(L)}$,矩阵元素为 $\hat{\theta}_{ii}$(请勿与全体估计参数的向量 $\hat{\boldsymbol{\theta}}$ 相混淆)。

在这个例子中,经验贝叶斯预测量可表示为

$$\tilde{\boldsymbol{\zeta}}_{z(L)}^{\mathrm{EB}} = \hat{\boldsymbol{\Psi}}_{(L)}\hat{\boldsymbol{\Lambda}}_{z(L)}'\hat{\boldsymbol{\Sigma}}_{z(L)}^{-1}(\mathbf{y}_{z(L)} - \mathbf{X}_{z(L)}\hat{\boldsymbol{\beta}}) \tag{7.3}$$

其中

$$\hat{\boldsymbol{\Sigma}}_{z(L)} \equiv \hat{\boldsymbol{\Lambda}}_{z(L)}\hat{\boldsymbol{\Psi}}_{(L)}\hat{\boldsymbol{\Lambda}}_{z(L)}' + \hat{\boldsymbol{\Theta}}_{z(L)}$$

即为 $\mathbf{y}_{z(L)}$ 的估计的残差协方差结构。对于因子模型,式(7.3)中的项 $\hat{\boldsymbol{\Psi}}_{(L)}\hat{\boldsymbol{\Lambda}}_{z(L)}'\hat{\boldsymbol{\Sigma}}_{z(L)}^{-1}$ 通常被称为该回归方法的**因子计分矩阵**(factor scoring matrix)。

由于式(7.3)中,处于括号内公式项的期望值为 0,因此(关于 $y_{z(L)}$ 的)非条件期望值就为

$$\mathrm{E}_{\mathbf{y}}(\tilde{\boldsymbol{\zeta}}_{z(L)}^{\mathrm{EB}} \mid \mathbf{X}_{z(L)};\hat{\boldsymbol{\theta}}) = \mathbf{0}$$

如果我们控制真实的潜变量 $\zeta_{z(L)}$,那么就不再满足

$$\mathrm{E}_{\mathbf{y}}(\mathbf{y}_{z(L)} - \mathbf{X}_{z(L)}\hat{\boldsymbol{\beta}} \mid \boldsymbol{\zeta}_{z(L)}, \mathbf{X}_{z(L)}; \hat{\boldsymbol{\theta}}) = \hat{\boldsymbol{\Lambda}}_{z(L)}\boldsymbol{\zeta}_{z(L)}$$

因而,对于给定 $\zeta_{z(L)}$ 所得到的条件期望就为

$$\begin{aligned}\mathrm{E}_{\mathbf{y}}(\tilde{\boldsymbol{\zeta}}^{\mathrm{EB}}_{z(L)} \mid \boldsymbol{\zeta}_{z(L)}, \mathbf{X}_{z(L)}; \hat{\boldsymbol{\theta}}) &= (\hat{\boldsymbol{\Psi}}_{(L)}\hat{\boldsymbol{\Lambda}}'_{z(L)}\hat{\boldsymbol{\Sigma}}^{-1}_{z(L)}\hat{\boldsymbol{\Lambda}}_{z(L)})\boldsymbol{\zeta}_{z(L)} \\ &= \boldsymbol{\zeta}_{z(L)} - (\mathbf{I} + \hat{\boldsymbol{\Psi}}_{(L)}\hat{\boldsymbol{\Omega}}_{z(L)})^{-1}\boldsymbol{\zeta}_{z(L)} \qquad (7.4)\end{aligned}$$

其中

$$\hat{\boldsymbol{\Omega}}_{z(L)} \equiv \hat{\boldsymbol{\Lambda}}'_{z(L)}\hat{\boldsymbol{\Theta}}^{-1}_{z(L)}\hat{\boldsymbol{\Lambda}}_{z(L)}$$

由于式(7.4)中最后一项通常不等于零向量,因而该经验贝叶斯预测量是非条件无偏的,而不是条件有偏的。在这个“线性案例”中,如果参数被当做已知(另见 Lawley and Maxwell,1971),则它使得预测误差的边缘样本方差最小化,那么经验贝叶斯预测量在 $\mathbf{y}_{(L)}$ 上是线性的,非条件无偏的,且为最优的估计,因此,该经验贝叶斯预测量为“最佳线性无偏预测量”(BLUP)。请注意,这个 BLUP 概念比参数经验贝叶斯更加一般化,因为它不依赖于对分布的假设(如 McCulloch and Searle,2001:256)。

事实上,统计学和计量心理学文献中的很多结果可被当做是从上述公式推导出来的特殊例子。对于常规的公共因子模型(见 3.3.2 节及 3.3.3 节),经验贝叶斯预测仅仅是汤普逊(Thomson,1938)以及沙斯通(Thurstone,1935)所讨论的因子计分的**回归方法**。有意思的是,斯皮尔曼(Spearman,1927)在他的单因子模型中所提出的估计量也是上面的一个特殊例子,也即是经验贝叶斯方法论的一个早期应用。

对于随机效应模型,我们得到劳(Rao,1975)以及斯特尼奥等(Strenio et al.,1983)所报告的结果。例如,在一个第一层次为同方差,$\theta_{ii} = \theta$ 的双层次随机截距模型中,经验贝叶斯预测量简化为

$$\begin{aligned}\tilde{\zeta}^{\mathrm{EB}}_{j(2)} &= \frac{\hat{\psi}}{\hat{\psi} + \hat{\theta}/n_j}\left(\frac{1}{n_j}\sum_{i=1}^{n_j}(y_{ij} - \mathbf{x}'_{ij}\hat{\boldsymbol{\beta}})\right) \\ &= \hat{R}_{j(2)}\left(\frac{1}{n_j}\sum_{i=1}^{n_j}(y_{ij} - \mathbf{x}'_{ij}\hat{\boldsymbol{\beta}})\right) \qquad (7.5)\end{aligned}$$

式中,括号中的项为群组 $j$ 的总残差或“原始”残差的均值,而 $\hat{R}_{j(2)}$ 为将经验贝叶斯预测抑制在朝向 0 的“**收缩**”因子,即先验分布的均值。该收缩因子可被理解为原始残差均值作为对 $\zeta_{j(2)}$ 的“测量”(“真实得分”的方差对总方差)的估计信度。当 $n_j$ 很小且 $\theta$ 相比 $\psi$ 很大时,该信度值最小;响应的条件密度 $\prod g^{(1)}(y_{ij} \mid \zeta_{j(2)}, \mathbf{x}_{ij}; \hat{\boldsymbol{\theta}})$ 变得很平,相比起先验密度 $h(\zeta_{j(2)}; \hat{\boldsymbol{\theta}})$ 并不提供多少信息。

在经验贝叶斯预测中,对小群组的先验的效应是将预测值 $\mathbf{x}'_{ij}\hat{\boldsymbol{\beta}} + \tilde{\zeta}_{j(2)}$ 拉向 $\mathbf{x}'_{ij}\hat{\boldsymbol{\beta}}$(其中所有群组都对 $\boldsymbol{\beta}$ 的估计有贡献),常被称为从其他群组“**借力**”。我们将在 7.6 节中证明收缩的概念在一般的“线性案例”中同样适用于经验贝叶斯预测。

对于一个三层次线性随机截距模型来说,用一个具体的例子是非常有帮助的,我们将不断使用本章的这个例子。

例:第三层次单元包含两个第二层次单元,且每个第二层次单元中包含两个第一层次单元。正如 3.2 节的图 3.2 所示,该结构矩阵 $\hat{\boldsymbol{\Lambda}}_{1(3)}$(在图中被称为 $\hat{\mathbf{Z}}_{1(3)}$)以及潜变量向量 $\boldsymbol{\zeta}_{1(3)}$ 为

$$\hat{\mathbf{\Lambda}}_{1(3)} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} \qquad \zeta_{1(3)} = \begin{bmatrix} \zeta_{11}^{(2)} \\ \zeta_{21}^{(2)} \\ \zeta_{11}^{(3)} \end{bmatrix}$$

假设随机效应协方差矩阵被估计为

$$\hat{\mathbf{\Psi}}_{(3)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

且第一层次方差为 $\hat{\theta} = 1$，那么经验贝叶斯预测则为

$$\tilde{\zeta}_{z(L)}^{\mathrm{EB}} = \begin{bmatrix} 0.21 & 0.21 & -0.12 & -0.12 \\ -0.12 & -0.12 & 0.21 & 0.21 \\ 0.18 & 0.18 & 0.18 & 0.18 \end{bmatrix} \begin{bmatrix} y_{11} - \mathbf{x}_{11}'\hat{\boldsymbol{\beta}} \\ y_{21} - \mathbf{x}_{21}'\hat{\boldsymbol{\beta}} \\ y_{31} - \mathbf{x}_{31}'\hat{\boldsymbol{\beta}} \\ y_{41} - \mathbf{x}_{41}'\hat{\boldsymbol{\beta}} \end{bmatrix}$$

其中，“权重”取两位小数点。请注意，在相同的顶层单元内的所有响应均对该单元提供关于潜变量的全部信息。给定**真实实现的潜变量**，则条件期望为

$$\mathrm{E}_{\mathbf{y}}(\tilde{\boldsymbol{\zeta}}_{z(L)}^{\mathrm{EB}} \mid \boldsymbol{\zeta}_{z(L)}) = \begin{bmatrix} 0.42 & -0.24 & 0.18 \\ -0.24 & 0.42 & 0.18 \\ 0.36 & 0.36 & 0.73 \end{bmatrix} \begin{bmatrix} \zeta_{11}^{(2)} \\ \zeta_{21}^{(2)} \\ \zeta_{11}^{(3)} \end{bmatrix}$$

在此，每个潜变量的真实实现影响在同一群组中的**所有**潜变量的预测均值。例如，平均来说，较大的第三层次真实随机截距数值将会导致对第二层次随机截距的过高估计。

值得注意的是，一般来说，潜得分不能简单地被代入潜变量的非线性函数以获得对该函数的估计。例如，在一个二项随机截距对数概率模型中，正数响应的概率的表达式为$[1+\exp(-\mathbf{x}_{ij}'\boldsymbol{\beta}-\zeta_j)]^{-1}$。为了得到该概率的经验贝叶斯预测，我们必须在考虑潜变量的后验分布的条件下对非线性函数求积分，而非代入$\tilde{\zeta}_j$。

## 7.3.2 经验贝叶斯方差以及协方差

我们分析 4 种类型的潜得分方差和协方差（我们将使用“协方差”来指称方差和协方差）：

- 后验协方差：$\mathrm{Cov}(\boldsymbol{\zeta}_{(L)} \mid \mathbf{y}_{(L)}, \mathbf{X}_{(L)}, \hat{\boldsymbol{\theta}})$
- 边缘抽样协方差：$\mathrm{Cov}_{\mathbf{y}}(\tilde{\boldsymbol{\zeta}}_{(L)} \mid \mathbf{X}_{(L)}, \hat{\boldsymbol{\theta}})$
- 条件抽样协方差：$\mathrm{Cov}_{\mathbf{y}}(\tilde{\boldsymbol{\zeta}}_{(L)} \mid \boldsymbol{\zeta}_{(L)}, \mathbf{X}_{(L)}, \hat{\boldsymbol{\theta}})$
- 预测误差协方差（边缘）：$\mathrm{Cov}_{\mathbf{y}}(\tilde{\boldsymbol{\zeta}}_{(L)} - \boldsymbol{\zeta}_{(L)} \mid \mathbf{X}_{(L)}, \hat{\boldsymbol{\theta}})$

这些协方差与本章所讨论的所有计分方法有关。对于结构参数 $\boldsymbol{\theta}$，我们用估计量 $\hat{\boldsymbol{\theta}}$ 替代。请注意，后验协方差不是关于随机参数向量 $\boldsymbol{\theta}$ 的边缘，因此它不是完全贝叶斯；由于 $\hat{\boldsymbol{\theta}}$ 的抽样变差被忽略了，抽样协方差并不属于**完全频率学派**。

**后验方差和协方差**

**后验**协方差矩阵是在给定观测响应和协变量条件下,潜变量在后验分布上的协方差矩阵。我们所依赖的**经验后验**协方差矩阵是将参数估计代入的后验协方差矩阵。基于后验均值以及后验标准差的置信区间与基于后验近似正态的贝叶斯可信区间是类似的。

经验后验协方差矩阵可用下式给出

$$\mathrm{Cov}(\boldsymbol{\zeta}_{(L)} \mid \mathbf{y}_{(L)}, \mathbf{X}_{(L)}; \hat{\boldsymbol{\theta}}) = \int (\boldsymbol{\zeta}_{(L)} - \tilde{\boldsymbol{\zeta}}_{(L)}^{\mathrm{EB}})(\boldsymbol{\zeta}_{(L)} - \tilde{\boldsymbol{\zeta}}_{(L)}^{\mathrm{EB}})' \omega(\boldsymbol{\zeta}_{(L)} \mid \mathbf{y}_{(L)}, \mathbf{X}_{(L)}; \hat{\boldsymbol{\theta}}) \mathrm{d}\boldsymbol{\zeta}_{(L)}$$

该公式也可通过 6.3.2 节的自适应积分法来进行数值积分而得到。

由于结构参数 $\boldsymbol{\theta}$ 被当做已知,因此与完全贝叶斯后验方差相比,经验后验方差向下偏移。通过将完全贝叶斯后验方差矩阵写为下述形式可看到这一点。

$$\mathrm{Cov}(\zeta_{(L)} \mid y_{(L)}, \mathbf{X}_{(L)}) = \mathrm{E}_{\boldsymbol{\theta}}[\mathrm{Cov}(\zeta_{(L)} \mid y_{(L)}, \mathbf{X}_{(L)}; \boldsymbol{\theta})] + \mathrm{Cov}_{\boldsymbol{\theta}}[\mathrm{E}(\zeta_{(L)} \mid y_{(L)}, \mathbf{X}_{(L)}; \boldsymbol{\theta})] \tag{7.6}$$

其中,第一项为经验贝叶斯后验方差矩阵的近似。重要的是,当群组个数较多且每个群组内单元个数少时(Kass and Steffey,1989),式(7.6)的第一项将起决定作用。另外,卡斯和斯特菲(Kass and Steffey,1989)还建议对于第二项也采用近似法。假设对于全部参数都存在单一先验,它们的一阶近似(用估计的基本参数 $\hat{\boldsymbol{\vartheta}}$ 表示)则为

$$\mathrm{Cov}_{\boldsymbol{\vartheta}}[\mathrm{E}(\boldsymbol{\zeta}_{(L)} \mid \mathbf{y}_{(L)}, \mathbf{X}_{(L)}; \hat{\boldsymbol{\vartheta}})] \approx -\mathbf{H}^{-1}\left(\frac{\partial \mathbf{E}(\boldsymbol{\zeta}_{(L)} \mid \mathbf{y}_{(L)}, \mathbf{X}_{(L)}; \hat{\boldsymbol{\vartheta}})}{\partial \boldsymbol{\vartheta}}\right)\left(\frac{\partial \mathrm{E}(\boldsymbol{\zeta}_{(L)} \mid \mathbf{y}_{(L)}, \mathbf{X}_{(L)}; \hat{\boldsymbol{\vartheta}})}{\partial \boldsymbol{\vartheta}}\right)'$$

其中,H 为 $\hat{\boldsymbol{\vartheta}}$ 处的对数似然海森矩阵,括号中的项为关于 $\boldsymbol{\vartheta}$ 的经验贝叶斯预测偏导数,$\boldsymbol{\vartheta}$ 用估计值 $\hat{\boldsymbol{\vartheta}}$ 计算。在线性混合模型中,该近似法存在闭式,戈德斯坦恩(Goldstein,2003,附录 2.2)讨论了这一点。对于混合效应对数概率回归,滕哈弗和罗加里奥(Ten Have and Localio,1999)使用数值积分来求解该近似。请见 9.5 节关于研究特定的试验效应的元分析模型中,对完整贝叶斯以及经验贝叶斯可信区间的比较。

后验标准差经常被频率学派当做经验贝叶斯预测的标准误差。在"线性案例"中,由于后验标准差等于预测误差的抽样标准差(见 7.3.1 节),因此是可以的。但是,后验标准差也常用于 IRT(如 Embretson and Reise,2000)以及广义线性混合模型(如 Ten Have and Localio,1999)中,很明显,在这两个模型中得不到频率学派的支持。

**"线性案例"**:在该情形下,经验后验协方差矩阵则成为

$$\mathrm{Cov}(\boldsymbol{\zeta}_{z(L)} \mid \mathbf{y}_{z(L)}, \mathbf{X}_{z(L)}; \hat{\boldsymbol{\theta}}) = \hat{\boldsymbol{\Psi}}_{(L)} - \hat{\boldsymbol{\Psi}}'_{(L)} \hat{\boldsymbol{\Lambda}}'_{z(L)} \hat{\boldsymbol{\Sigma}}_{z(L)}^{-1} \hat{\boldsymbol{\Lambda}}_{z(L)} \hat{\boldsymbol{\Psi}}_{(L)} \tag{7.7}$$

对于这个特定的随机截距模型来说,则简化为

$$\mathrm{Var}(\zeta_{j(2)} \mid \mathbf{y}_{j(2)}, \mathbf{X}_{j(2)}; \hat{\boldsymbol{\theta}}) = \frac{\hat{\theta}/n_j}{\hat{\psi} + \hat{\theta}/n_j}\hat{\psi} = (1 - \hat{R}_{j(2)})\hat{\psi}$$

正如事先所料,由于知道了响应 $\mathbf{y}_{j(2)}$,在随机截距上获得了更多信息,因此,后验方差小于先验方差。

**例**:重新回到前面那个三层次模型的例子,其后验协方差矩阵成为

$$\mathrm{Cov}(\boldsymbol{\zeta}_{1(3)} \mid \mathbf{y}_{1(3)}, \mathbf{X}_{1(3)}; \hat{\boldsymbol{\theta}}) = \begin{bmatrix} 0.58 & & \\ 0.24 & 0.58 & \\ -0.36 & -0.36 & 0.55 \end{bmatrix}$$

有两个重要的地方需要指出。首先,虽然在先验分布下,那两个第二层次单元的随机截距不相关,但是后验协方差并不为0(等于0.24)。其次,尽管在先验分布下,不存在跨层次的协方差,但是第二层次的随机截距与第三层次随机截距的协方差并不等于0(等于-0.36)。

**边缘抽样方差和协方差**

**边缘抽样**协方差为在固定协变量以及参数估计的条件下,对群组和群组内部单元进行重复抽样所得到的估计量的协方差。

与下一节所讨论的条件抽样协方差不同,边缘抽样协方差反映了由于抽样所造成的潜变量与它们的先验分布之间的变差。因此,边缘抽样标准差就可用于检测模型中不一致的那些群组。由于这个原因,戈德斯坦恩(Goldstein,2003)将此特性称为"**诊断性标准误**(diagnostic standard error)"。8.6.2 节对此做了进一步讨论。

经验贝叶斯预测量的边缘抽样协方差矩阵为

$$\begin{aligned}\mathrm{Cov}_{\mathrm{y}}(\tilde{\boldsymbol{\zeta}}_{(L)}^{\mathrm{EB}} \mid \mathbf{X}_{(L)};\hat{\boldsymbol{\theta}}) &= \mathrm{Cov}_{\mathrm{y}}[\mathrm{E}(\boldsymbol{\zeta}_{(L)} \mid \mathbf{y}_{(L)},\mathbf{X}_{(L)};\hat{\boldsymbol{\theta}})] \\ &= \int \tilde{\boldsymbol{\zeta}}_{(L)}^{\mathrm{EB}}\tilde{\boldsymbol{\zeta}}_{(L)}^{\mathrm{EB}\,\prime} g^{(L)}(\mathbf{y}_{(L)} \mid \mathbf{X}_{(L)};\hat{\boldsymbol{\theta}})\,\mathrm{d}\mathbf{y}_{(L)}\end{aligned}$$

其中,$g^{(L)}(\mathbf{y}_{(L)} \mid \mathbf{X}_{(L)};\hat{\boldsymbol{\theta}})$ 为处于顶层群组的响应的联合边缘分布。请注意,为了标记简单,我们忽略了下标 $z$ 并在本章的随后部分都做如此处理。

不幸的是,不存在一般模型的闭式表达式。

然而,斯科隆多(Skrondal,1996)提出根据关系

$$\begin{aligned}\mathrm{Cov}(\boldsymbol{\zeta}_{(L)} \mid \mathbf{X}_{(L)};\hat{\boldsymbol{\theta}}) =& \mathrm{E}_{\mathrm{y}}[\mathrm{Cov}(\boldsymbol{\zeta}_{(L)} \mid \mathbf{y}_{(L)},\mathbf{X}_{(L)};\hat{\boldsymbol{\theta}})] + \\ & \mathrm{Cov}_{\mathrm{y}}[\mathrm{E}(\boldsymbol{\zeta}_{(L)} \mid \mathbf{y}_{(L)},\mathbf{X}_{(L)};\hat{\boldsymbol{\theta}})]\end{aligned}$$

来推导出一个近似的表达式。由于最后一项为经验贝叶斯预测量必需的协方差矩阵,该式可重写为

$$\mathrm{Cov}_{\mathrm{y}}(\tilde{\boldsymbol{\zeta}}_{(L)}^{\mathrm{EB}} \mid \mathbf{X}_{(L)};\hat{\boldsymbol{\theta}}) = \mathrm{Cov}(\boldsymbol{\zeta}_{(L)} \mid \mathbf{X}_{(L)};\hat{\boldsymbol{\theta}}) - \mathrm{E}_{\mathrm{y}}[\mathrm{Cov}(\boldsymbol{\zeta}_{(L)} \mid \mathbf{y}_{(L)},\mathbf{X}_{(L)};\hat{\boldsymbol{\theta}})]$$

在该式中,右边部分的第一项为潜变量 $\boldsymbol{\zeta}_{(L)}$ 向量的估计先验协方差矩阵 $\hat{\boldsymbol{\Psi}}$,第二项为后验协方差矩阵的期望值,并可用后验协方差矩阵来近似逼近。因而,我们得到

$$\mathrm{Cov}_{\mathrm{y}}(\tilde{\boldsymbol{\zeta}}_{(L)}^{\mathrm{EB}} \mid \mathbf{X}_{(L)};\hat{\boldsymbol{\theta}}) \approx \hat{\Psi}_{(L)} - \mathrm{Cov}(\boldsymbol{\zeta}_{(L)} \mid \mathbf{y}_{(L)},\mathbf{X}_{(L)};\hat{\boldsymbol{\theta}}) \tag{7.8}$$

另外一种方法为通过模拟来获得近似,首先从先验分布中对潜变量进行抽样,然后在给定潜变量和协变量的条件下从条件分布中对响应进行抽样。这个方法的优点在于,存在于参数估计 $\hat{\boldsymbol{\theta}}$ 中的不确定性可很容易地通过在对潜变量抽样之前,从抽样分布中抽取一个新的样本来解决;朗福特(Longford,2001)在一个类似的情形中提出了该方法。

"**线性案例**":在这个情形中,式(7.8)完全成立,边缘抽样方差成为

$$\begin{aligned}\mathrm{Cov}_{\mathrm{y}}(\tilde{\boldsymbol{\zeta}}_{(L)}^{\mathrm{EB}} \mid \mathbf{X}_{(L)};\hat{\boldsymbol{\theta}}) &= \hat{\boldsymbol{\Psi}}_{(L)} - \mathrm{Cov}(\boldsymbol{\zeta}_{(L)} \mid \mathbf{y}_{(L)},\mathbf{X}_{(L)};\hat{\boldsymbol{\theta}}) \\ &= \hat{\boldsymbol{\Psi}}_{(L)}\hat{\boldsymbol{\Lambda}}_{(L)}^{\prime}\hat{\boldsymbol{\Sigma}}_{(L)}^{-1}\hat{\boldsymbol{\Lambda}}_{(L)}\hat{\boldsymbol{\Psi}}_{(L)}\end{aligned} \tag{7.9}$$

由于后验方差为正数,该协方差矩阵对角线上的元素比先验协方差矩阵上的明显小得多。这是由于收缩造成的。

由于收缩,一些研究者(如 Louis,1984)提出使用和先验分布具有相同协方差的**修正经验贝叶斯预测量**(adjusted empirical Bayes predictors)。该预测量可使得式(7.2)的(对于给定参数估计的)加总二次损失函数的后验期望值最小化,该损失函数所服从的边际

条件为预测值满足先验分布的一阶矩和二阶矩估计。在因子分析文献中,根据与先验分布相同的协方差矩阵来获得因子得分的想法最早来自安德森和鲁宾(Anderson and Rubin,1956),他们对正交因子模型进行了研究。这种"保留协方差"的方法已被扩展到一般先验协方差矩阵中(如 Ten Berge,1983;Ten Berge et al.,1999)。

对于简单随机截距模型,边缘抽样方差就简化为

$$\mathrm{Var}_{\mathbf{y}}(\tilde{\zeta}_{j(2)}^{\mathrm{EB}} \mid \mathbf{X}_{j(2)};\hat{\boldsymbol{\theta}}) = \frac{\hat{\psi}}{\hat{\psi}+\hat{\theta}/n_j}\hat{\psi} = \hat{R}_{j(2)}\hat{\psi}$$

**例**:回到三层次随机截距模型例子上来,第三层次单元的随机效应的边缘抽样方差矩阵就成为

$$\mathrm{Cov}_{\mathbf{y}}(\tilde{\boldsymbol{\zeta}}_{1(3)}^{\mathrm{EB}} \mid \mathbf{X}_{1(3)};\hat{\boldsymbol{\theta}}) = \begin{bmatrix} 0.42 & & \\ -0.24 & 0.42 & \\ 0.36 & 0.36 & 1.45 \end{bmatrix}$$

请注意,正如后验协方差矩阵那样,第二层次上不同单元的随机截距之间的协方差并不为0。另外,层次之间的协方差也不为0。

层次之间存在的不等于零的抽样协方差使得多层次模型的诊断变得相当复杂。不幸的是,在关于多层次模型的文献中,这个问题通常被忽视,如朗福特和路易斯(Langford and Lewis,1998)及(Goldstein,2003)。

**条件抽样方差和协方差**

**条件抽样**协方差为在固定的"真实"潜变量 $\zeta_{(L)}$(以及固定的协变量和参数估计)条件下,在同一群组中对单元进行重复抽样所得到的预测的协方差。

注意,请勿将条件抽样标准差与戈德斯坦恩(Goldstein,2003:23)所用的"比较标准误差(comparative standard error)"相混淆。戈德斯坦恩将它描述为在条件依赖于真实潜变量,但事实上,它是预测误差的**边缘**(对于$\boldsymbol{\zeta}$)抽样方差,在7.4.1节,我们将证明在"线性案例"中,它等于后验方差。

在给定潜变量的条件下,经验贝叶斯预测量 $\tilde{\boldsymbol{\zeta}}_{(L)}^{\mathrm{EB}}$ 的条件抽样协方差矩阵为

$$\begin{aligned}\mathrm{Cov}_{\mathbf{y}}(\tilde{\boldsymbol{\zeta}}_{(L)}^{\mathrm{EB}} \mid \boldsymbol{\zeta}_{(L)},\mathbf{X}_{(L)};\hat{\boldsymbol{\theta}}) &= \mathrm{Cov}_{\mathbf{y}}[E(\boldsymbol{\zeta}_{(L)} \mid \mathbf{y}_{(L)},\mathbf{X}_{(L)};\hat{\boldsymbol{\theta}}) \mid \boldsymbol{\zeta}_{(L)}] \\ &= \int \tilde{\zeta}_{(L)}^{\mathrm{EB}}\tilde{\zeta}_{(L)}^{\mathrm{EB}\prime}\prod g^{(1)}(y_{ij\cdots z} \mid \boldsymbol{\zeta}_{(L)},\mathrm{X}_{(L)};\hat{\boldsymbol{\theta}})\,\mathrm{d}\mathbf{y}_{(L)}\end{aligned}$$

其中,乘积表示在该 $L$ 层次单元内的所有第一层次单元的响应的联合条件分布。

**"线性案例"**:在此,条件抽样协方差矩阵成为

$$\mathrm{Cov}_{\mathbf{y}}(\tilde{\boldsymbol{\zeta}}_{(L)}^{\mathrm{EB}} \mid \boldsymbol{\zeta}_{(L)},\mathbf{X}_{(L)};\hat{\boldsymbol{\theta}}) = \hat{\boldsymbol{\Psi}}'_{(L)}\hat{\boldsymbol{\Lambda}}'_{(L)}\hat{\boldsymbol{\Sigma}}_{(L)}^{-1}\hat{\boldsymbol{\Theta}}_{(L)}\hat{\boldsymbol{\Sigma}}_{(L)}^{-1}\hat{\boldsymbol{\Lambda}}_{(L)}\hat{\boldsymbol{\Psi}}_{(L)} \tag{7.10}$$

对于随机截距模型,可表示为

$$\mathrm{Var}_{\mathbf{y}}(\tilde{\zeta}_{j(2)}^{\mathrm{EB}} \mid \zeta_{j(2)},\mathbf{X}_{j(2)};\hat{\boldsymbol{\theta}}) = \hat{R}_{j(2)}(1-\hat{R}_{j(2)})\hat{\psi}$$

**例**:在这个三层次随机截距模型中,第三层次单元的随机效应的条件抽样协方差矩阵就为

$$\mathrm{Cov}_{\mathbf{y}}(\tilde{\boldsymbol{\zeta}}_{1(3)}^{\mathrm{EB}} \mid \boldsymbol{\zeta}_{1(3)},\mathbf{X}_{1(3)};\hat{\boldsymbol{\theta}}) = \begin{bmatrix} 0.12 & & \\ -0.10 & 0.12 & \\ 0.03 & 0.03 & 0.13 \end{bmatrix}$$

正如所料,该方差明显小于边缘抽样方差。另外在这个例子中,请注意条件抽样方

差矩阵与所对应的后验协方差矩阵之间的区别，这些方差经常被混淆。

**预测误差方差和协方差**

**预测误差**协方差为根据响应的边缘分布对其进行重复抽样所得到的预测误差 $\tilde{\boldsymbol{\zeta}}_{(L)}^{\mathrm{EB}}-\boldsymbol{\zeta}_{(L)}$ 的协方差，即

$$\mathrm{Cov}_{\mathbf{y}}(\tilde{\boldsymbol{\zeta}}_{(L)}^{\mathrm{EB}}-\boldsymbol{\zeta}_{(L)} \mid \mathbf{X}_{(L)};\hat{\boldsymbol{\theta}})=\int(\tilde{\boldsymbol{\zeta}}_{(L)}^{\mathrm{EB}}-\boldsymbol{\zeta}_{(L)})(\tilde{\boldsymbol{\zeta}}_{(L)}^{\mathrm{EB}}-\boldsymbol{\zeta}_{(L)})'g^{(L)}(\mathbf{y}_{(L)} \mid \mathbf{X}_{(L)};\hat{\boldsymbol{\theta}})\,\mathrm{d}\mathbf{y}_{(L)}$$

该协方差是边缘的，因为它们反映对潜变量以及在给定潜变量条件下对响应进行抽样所产生的变差。预测误差的标准差可能是最明显关于预测不确定性的测量。

使用瓦科拉维和梁(Waclawiw and Liang,1994)所提出的**导数多元概括法**(multivariate generalization)，预测误差协方差矩阵可表示为(为了简短起见，忽略对 $\mathbf{X}_{(L)}$ 和 $\hat{\boldsymbol{\theta}}$ 的条件依赖)

$$\begin{aligned}
&\mathrm{Cov}_{\mathbf{y}}(\tilde{\boldsymbol{\zeta}}_{(L)}^{\mathrm{EB}}-\boldsymbol{\zeta}_{(L)})\\
&=\mathrm{E}_{\mathbf{y}}[(\tilde{\boldsymbol{\zeta}}_{(L)}^{\mathrm{EB}}-\boldsymbol{\zeta}_{(L)})(\tilde{\boldsymbol{\zeta}}_{(L)}^{\mathrm{EB}}-\boldsymbol{\zeta}_{(L)})']-\mathrm{E}_{\mathbf{y}}(\tilde{\boldsymbol{\zeta}}_{(L)}^{\mathrm{EB}}-\boldsymbol{\zeta}_{(L)})\mathrm{E}_{\mathbf{y}}(\tilde{\boldsymbol{\zeta}}_{(L)}^{\mathrm{EB}}-\boldsymbol{\zeta}_{(L)})'\\
&=\mathrm{E}_{\mathbf{y}}[(\tilde{\boldsymbol{\zeta}}_{(L)}^{\mathrm{EB}}-\boldsymbol{\zeta}_{(L)})(\tilde{\boldsymbol{\zeta}}_{(L)}^{\mathrm{EB}}-\boldsymbol{\zeta}_{(L)})']\\
&=\mathrm{E}_{\mathbf{y}}[(\mathrm{E}(\boldsymbol{\zeta}_{(L)} \mid \mathbf{y}_{(L)})-\boldsymbol{\zeta}_{(L)})(\mathrm{E}(\boldsymbol{\zeta}_{(L)} \mid \mathbf{y}_{(L)})-\boldsymbol{\zeta}_{(L)})']\\
&=\mathrm{E}_{\mathbf{y}}\{\mathrm{E}_{\mathbf{y}}[(\mathrm{E}(\boldsymbol{\zeta}_{(L)} \mid \mathbf{y}_{(L)})-\boldsymbol{\zeta}_{(L)})(\mathrm{E}(\boldsymbol{\zeta}_{(L)} \mid \mathbf{y}_{(L)})-\boldsymbol{\zeta}_{(L)})' \mid \mathbf{y}_{(L)}]\}\\
&=\mathrm{E}_{\mathbf{y}}[\mathrm{Cov}(\boldsymbol{\zeta}_{(L)} \mid \mathbf{y}_{(L)})]
\end{aligned}$$

即关于边缘抽样分布的期望后验协方差矩阵。在上面的推导中，第二个等式依赖于经验贝叶斯预测的非条件无偏性，第四个等式对**双重期望规则**进行展开，最后一个等式只是用于说明方括号中的项表示后验协方差矩阵。

上式表明存在如下的近似：

$$\mathrm{Cov}_{\mathbf{y}}(\tilde{\boldsymbol{\zeta}}_{(L)}^{\mathrm{EB}}-\boldsymbol{\zeta}_{(L)} \mid \mathbf{X}_{(L)};\hat{\boldsymbol{\theta}})\approx\mathrm{Cov}(\boldsymbol{\zeta}_{(L)} \mid \mathbf{y}_{(L)},\mathbf{X}_{(L)};\hat{\boldsymbol{\theta}}) \tag{7.11}$$

另一种方法为与边缘抽样协方差类似的模拟法：根据潜变量的先验抽取潜变量，然后在给定潜变量的条件下根据条件分布对响应变量进行抽样。“真实的”潜变量实现即为模拟出来的潜变量，将其从经验贝叶斯预测中减去，我们就可估计出预测误差方差。为了表明这种参数估计方法的非严密性，上述方法也可通过从所估计的抽样分布中对参数进行抽样来进行。

“线性案例”：式(7.7)中“线性案例”的后验协方差矩阵并不包含响应 $\mathbf{y}_{(L)}$，因此，在这个情形中，式(7.11)的近似就成为精确解，即

$$\begin{aligned}
\mathrm{Cov}_{\mathbf{y}}(\tilde{\boldsymbol{\zeta}}_{(L)}^{\mathrm{EB}}-\boldsymbol{\zeta}_{(L)} \mid \mathbf{X}_{(L)};\hat{\boldsymbol{\theta}})&=\hat{\boldsymbol{\Psi}}_{(L)}-\hat{\boldsymbol{\Psi}}'_{(L)}\hat{\boldsymbol{\Lambda}}'_{(L)}\hat{\boldsymbol{\Sigma}}_{z(L)}^{-1}\hat{\boldsymbol{\Lambda}}_{z(L)}\hat{\boldsymbol{\Psi}}_{(L)}\\
&=\mathrm{Cov}(\boldsymbol{\zeta}_{(L)} \mid \mathbf{y}_{(L)},\mathbf{X}_{(L)};\hat{\boldsymbol{\theta}})
\end{aligned}$$

即预测误差协方差就是式(7.7)所给出的后验协方差。

## 7.4 经验贝叶斯众数(Empirical Bayes modal,EBM)

### 7.4.1 经验贝叶斯众数预测

在经验贝叶斯预测中，我们也可使用后验众数，而非后验均值。可以证明，后验众数使得0—1损失函数的后验期望最小化：

$$L^{BM}(\boldsymbol{\zeta}_{(L)},\tilde{\boldsymbol{\zeta}}_{(L)}) = \begin{cases} 0 & 当 |\boldsymbol{\zeta}_{(L)} - \tilde{\boldsymbol{\zeta}}_{(L)}| \le \boldsymbol{\epsilon} \\ 1 & 当 |\boldsymbol{\zeta}_{(L)} - \tilde{\boldsymbol{\zeta}}_{(L)}| > \boldsymbol{\epsilon} \end{cases}$$

其中,$\boldsymbol{\epsilon}$为小数向量,使得当$\tilde{\boldsymbol{\zeta}}_{(L)}$接近于$\boldsymbol{\zeta}_{(L)}$时,$L^{BM}(\boldsymbol{\zeta}_{(L)},\tilde{\boldsymbol{\zeta}}_{(L)})$等于0,反之则等于1。与经验贝叶斯所暗含的均值平方误差损失函数不同,上述的损失函数对于潜类模型来说也是有意义的,因为当所预测的潜类不是真实潜类时,损失为1,反之为0。在这个情形下,损失函数仅仅是错误分类的个数。

对于$\theta$代入其估计值$\hat{\theta}$,可得

$$\tilde{\boldsymbol{\zeta}}_{(L)}^{EBM} = \underset{\boldsymbol{\zeta}_{(L)}}{\max\ \arg}\ \omega(\boldsymbol{\zeta}_{(L)} \mid \mathbf{y}_{(L)},\mathbf{X}_{(L)};\hat{\boldsymbol{\theta}})$$

我们建议将此预测量作为**经验贝叶斯众数**(EBM)。有些学者称之为贝叶斯众数(如Samejima,1969)或者"后验估计量众数(MAP)"(如 Bock and Aitkin,1981;Bock and Mislevy,1982;Bock,1983,1985),但这些名称都没能够表明估计值已被代入。

在潜类模型中,经验贝叶斯众数是标准的分类方法。另一个可选的损失函数对于不同的错误分类赋予不同的权重值,以反映不同的**代价**(cost)。贝叶斯风险标准采用使得期望代价最小化的分类方法。从另一个方面来说,**比例预测**(Proportional prediction)根据类别的后验概率将类别随机赋予不同的群组。克劳格等(Clogg et al.,1991)提出在多重插值中采用该方法对潜类进行插值计算。

一般来说,对于经验贝叶斯众数预测量来说,不存在解析式,我们必须求诸于数值法。后验众数是下式的解(在满足二阶条件的情况下)

$$\frac{\partial}{\partial \boldsymbol{\zeta}_{(L)}} \ln\ \omega(\boldsymbol{\zeta}_{(L)} \mid \mathbf{y}_{(L)},\mathbf{X}_{(L)};\hat{\boldsymbol{\theta}}) = 0$$

使用梯度法可获得上式(见6.4.2节),如牛顿-拉夫逊算法。从式(7.1)可见,后验分布的分母并不依赖于$\boldsymbol{\zeta}_{(L)}$,因此,在上述公式中,我们可以使用分子来代替$\omega$,即

$$\frac{\partial}{\partial \boldsymbol{\zeta}_{(L)}} \ln\ h(\boldsymbol{\zeta}_{(L)};\hat{\boldsymbol{\theta}}) + \frac{\partial}{\partial \boldsymbol{\zeta}_{(L)}} \ln \prod g^{(1)}(y_{ij\cdots z} \mid \mathbf{x}_{ij\cdots z},\boldsymbol{\zeta}_{(L)};\hat{\boldsymbol{\theta}}) = 0 \qquad (7.12)$$

与经验贝叶斯相反,该方法并不要求数值积分。由此,当后验密度为近似多元正态时,经验贝叶斯众数经常被当做是对经验贝叶斯的近似。在前一章提到过,林莱和斯密斯(Lindley and Smith,1972)提出在真实贝叶斯情形下,用贝叶斯众数来作为贝叶斯预测量的近似。使用该方法对应于在给定参数值条件下,使得关于潜变量的$h$-似然最大化的方法(见6.3.1节)。

由于避免了积分,对于包含非正态响应的潜变量模型来说,经验贝叶斯众数比经验贝叶斯在计算上效率更高。因此,萨姆基玛(Samejima,1969)将经验贝叶斯众数引入到她的"评分响应模型(graded response model)",即定序正态概率或包含正态分布因子的对数概率单因子模型中,并给出了关于其属性的严格推导。有趣的是,萨姆基玛报告了一个比较令人惊讶的结果,她发现对于一个只包含6个二项响应的模型来说,无法区分EB预测和EBM预测。穆森(Muthén,1977)将萨姆基玛的方法扩展到包含二项响应的正态概率多因子模型中。根据萨姆基玛和穆森的结果,斯科隆多(Skrondal,1996,第7章)讨论了包含多元正态潜在响应的多维潜变量模型的经验贝叶斯众数预测。

**"线性案例"**:在这里,后验为多元正态,使得期望值等于众数。因此,经验贝叶斯预测量与经验贝叶斯众数预测量相一致,见式(7.3)。

### 7.4.2 经验贝叶斯众数协方差与分类误差

**经验贝叶斯众数方差与协方差**

对于一般潜变量模型，我们并不清楚用于推导出经验贝叶斯众数的不同类型的抽样方差(协方差)的方法。然而，对于大的群组来说，其后验趋向于正态且后验均值接近于众数，经验贝叶斯抽样协方差应该是较好的近似。在“线性案例”中，经验贝叶斯协方差等于经验贝叶斯众数协方差；关于预测误差协方差，见式(7.7)，关于边缘抽样协方差见式(7.9)，对于条件抽样协方差见式(7.10)。

为了找到后验众数，使用了梯度法，很自然这里就会采用位于众数的海森矩阵的负逆矩阵(negative inverse of the Hessian)作为后验协方差矩阵的近似，即

$$\mathrm{Cov}(\zeta_{z(L)} \mid \mathbf{y}_{z(L)}, \mathbf{X}_{z(L)}; \hat{\theta}) \approx -\left(\frac{\partial^2}{\partial \zeta^2} \ln \omega(\zeta_{(L)} \mid \mathbf{y}_{(L)}, \mathbf{X}_{(L)}; \hat{\boldsymbol{\theta}})\right)^{-1}$$

随着后验逼近于多元正态，也就是随着群组规模增大，近似趋近精确。

**分类误差**

对于离散潜变量，给定群组的分类质量可用在给定 $\mathbf{y}_j$ 与 $\mathbf{X}_j$ 条件下，所估计的错误分类条件概率来评估，即

$$f_j = 1 - \max_c \omega(e_c \mid \mathbf{y}_j, \mathbf{X}_j; \hat{\boldsymbol{\theta}}) \qquad (7.13)$$

总体的分类错误率可用 $f_j$ 对群组求样本均值来估计。这也是成比例削减分类误差标准的基础，该标准将此分类错误率与在 $\mathbf{y}_j$ 与 $\mathbf{X}_j$ 都是未知条件下的分类错误率相比较(见 9.3 节和 13.5 节的应用)。

## 7.5 最大似然法

### 7.5.1 潜得分估计

潜变量有时被认为是非随机的或固定的。在这种情形下，很自然地会将潜得分理解为需要估计的未知参数。在 6.10.1 节中，我们分析了模型参数和潜变量的联合估计。与其相反，在此我们假设模型参数已被估计(采用前一章所讨论的某种估计方法)并考虑在给定 $\hat{\boldsymbol{\theta}}$ 条件下的潜变量 $\boldsymbol{\zeta}$ 的估计。

计分或分类的估计方法基于给定潜变量条件下的响应的条件分布，并代入模型参数的估计值 $\hat{\boldsymbol{\theta}}$，即

$$\prod g^{(1)}(y_{ij\cdots z} \mid \mathbf{x}_{ij\cdots z}, \zeta_{z(L)}; \hat{\boldsymbol{\theta}})$$

该条件分布被解释为将群组的潜变量数值当做未知参数的“似然”。

与模型参数的最大似然估计方法类似，通过解下述似然方程，该条件分布为关于未知潜变量的最大化，即

$$\frac{\partial}{\partial \zeta_{z(L)}} \ln \prod g^{(1)}(y_{ij\cdots z} \mid \mathbf{x}_{ij\cdots z}, \zeta_{z(L)}; \hat{\boldsymbol{\theta}}) = \mathbf{0}$$

值得指出的是，该公式相当于在经验贝叶斯众数中使式(7.12)的第二项最大化。在经验贝叶斯众数中，对数先验密度有可能因此被认为是先验众数偏差的**惩罚项**(penalty term)。因此，相对于最大似然估计来说，经验贝叶斯众数预测为朝向先验众数的收缩。

正如所料,随着群组中单元的个数趋向于无穷,对群组的估计值为渐进无偏。然而,这个结果可能毫无用处,因为群组中单元的个数通常都很少,如纵贯性研究或家庭研究。对于拉什模型,霍伊廷克和波姆斯玛(Hoijtink and Boomsma,1995)对关于最大似然、经验贝叶斯众数以及沃姆(Warm,1989)所提出的加权最大似然估计量的有限样本属性的文献进行了综述。

因为潜变量的先验分布并未被利用,稀疏信息群组还存在特别的问题。例如,请考虑一个包含二项响应的单维因子模型。如果对于一个给定群组(代表主体)的全部响应都为0,对于该群组的似然贡献无法达到最大化,为了满足似然方程,因子得分就成为$-\infty$(Samejima,1969)。另外一个例子为包含随机截距和斜率的增长曲线模型,如果对于一个群组只有一个观测响应,那么该群组的斜率就无法被估计。对于利用了先验分布的经验贝叶斯预测来说,这两个例子都不成为问题。第一个例子受益于收缩,将预测从$-\infty$牵引开,而第二个例子则受益于随机截距所包含的信息,该信息来自于其与随机斜率的后验协方差。

此外,潜变量的最大似然估计要求潜变量被当做固定参数。这与我们的模型框架以及参数估计的边缘最大似然方法相矛盾。基于上述理由,一般来说,我们不建议采用最大似然计分法。然而,在评估潜变量的正态假设时,最大似然方法可能也是有用的(见8.6.2节)。

"线性案例":解$\boldsymbol{\zeta}_{(L)}$的似然方程,得

$$\tilde{\boldsymbol{\zeta}}_{(L)}^{\mathrm{ML}} = (\hat{\boldsymbol{\Lambda}}'_{(L)}\hat{\boldsymbol{\Theta}}_{(L)}^{-1}\hat{\boldsymbol{\Lambda}}_{(L)})^{-1}\hat{\boldsymbol{\Lambda}}'_{(L)}\hat{\boldsymbol{\Theta}}_{(L)}^{-1}(\mathbf{y}_{(L)} - \mathbf{X}_{(L)}\boldsymbol{\beta}) \tag{7.14}$$

由此可得

$$\mathrm{E}_{\mathrm{y}}(\tilde{\boldsymbol{\zeta}}_{(L)}^{\mathrm{ML}} \mid \boldsymbol{\zeta}_{(L)}, \mathbf{X}_{(L)}; \hat{\boldsymbol{\theta}}) = \boldsymbol{\zeta}_{(L)}$$

因此,与经验贝叶斯以及贝叶斯众数预测量不同,在给定潜变量$\boldsymbol{\zeta}_{(L)}$的数值的条件下,最大似然估计量为**条件无偏**。

对于公共因子模型(见3.3.2节及3.3.3节),最大似然对应于**巴特利特**因子计分法(Bartlett,1937;1938)。在常规处理中,巴特利特方法被推导为使得标准残差的总方差最小化(Anderson and Rubin,1956;Lawley and Maxwell,1971)或者使得条件无偏估计量的加总二次损失最小化(Lawley and Maxwell,1971),因此,对于"线性案例"来说,无法将该方法理解为最大似然估计量。从另一个方面来说,这些推导恰恰说明使用巴特利特方法并不需要做出分布假设。

在随机效应模型中,该估计量也被当做随机效应的普通最小二乘法(OLS)估计量。在这个双层次随机截距模型的特殊例子中,该估计量就是原始残差均值

$$\tilde{\zeta}_{j(2)}^{\mathrm{ML}} = \frac{1}{n_j}\sum_{i=1}^{n_j}(y_{ij} - \mathbf{x}'_{ij}\hat{\boldsymbol{\beta}})$$

请注意,如果没有设定约束,在多个层次上包含固定截距的多层次模型不可辨识,例如,在同一个第三层次群组内的第二层次截距的总和为零。这说明对于本章的这个例子,最大似然估计并不存在,因此,式(7.14)的矩阵$(\hat{\boldsymbol{\Lambda}}'_{(L)}\hat{\boldsymbol{\Theta}}_{(L)}^{-1}\hat{\boldsymbol{\Lambda}}_{(L)})^{-1}$为奇异矩阵。

### 7.5.2 ML 估计量的方差与协方差

根据似然理论,ML 估计量$\tilde{\boldsymbol{\zeta}}_{(L)}^{\mathrm{ML}}$的近似协方差矩阵为

$$\mathrm{Cov}_y(\tilde{\boldsymbol{\zeta}}_{(L)}^{\mathrm{ML}} \mid \boldsymbol{\zeta}_{(L)}, \mathbf{X}_{(L)}; \hat{\boldsymbol{\theta}}) \approx -\left(\frac{\partial^2}{\partial\boldsymbol{\zeta}_{(L)}^2}\ln\prod g^{(1)}(y_{ij\cdots z} \mid \mathbf{x}_{ij\cdots z}, \boldsymbol{\zeta}_{(L)}; \hat{\boldsymbol{\theta}})\right)^{-1}$$

令人感兴趣的是，$\mathrm{Cov}_y(\tilde{\boldsymbol{\zeta}}_{(L)}^{\mathrm{ML}} | \boldsymbol{\zeta}_{(L)} \mathbf{X}_{z(L)};\hat{\boldsymbol{\theta}})$ 可被理解为在给定真实潜变量 $\boldsymbol{\zeta}_{z(L)}$ 的条件下，得分的**条件抽样**协方差矩阵。

该近似要求在群组中的单元（或题器）个数趋向于无穷，而非群组的个数。在实践中，每个群组的单元个数通常很少，因此，对这个结果的利用是存疑的。对于给定群组，上述协方差矩阵（即观测到的信息）的逆矩阵为个体单元的贡献 $-\partial^2 \ln g^{(1)}(y_{ij\cdots z} | \mathbf{x}_{ij\cdots z}, \boldsymbol{\zeta}_{(L)};\hat{\boldsymbol{\theta}})/\partial \boldsymbol{\zeta}_{(L)}^2$ 的总和。在单维 IRT（$\boldsymbol{\zeta}_{(L)} = \boldsymbol{\zeta}_{(j)}$）中，将这些个体题器的贡献对应于 $\zeta_{(j)}$ 作图，被称为"题器信息曲线"，所有题器的贡献的总和被称为"测试信息"（如 Birnbaum，1968）。提起信息函数可用于评估在得知被测者对一个题器的响应并将此作为其真实能力的函数之后，获得了多少关于该被测者能力的信息。在 IRT 中，信息函数有着非常重要的作用，尤其是在构念检验、题器选择、测量精度评估、检验比较、计分方法比较及适应性测试中（如 Hambleton and Swaminathan，1985）。例如，在适应性测试中，当前信息函数可根据对被测者能力的当前估计，选择下一个测试题器。如果题器对于被测者来说过于简单或者太难，那么该题器就无法提供信息，从而使得无法预测响应。

可以看到，随着群组规模趋向于无穷，最大似然估计量的协方差矩阵趋向于后验协方差矩阵，即似然落入先验（如 DeGroot，1970）。

边缘抽样协方差矩阵为

$$\mathrm{Cov}_y(\tilde{\boldsymbol{\zeta}}_{(L)}^{\mathrm{ML}} | \mathbf{X}_{(L)};\hat{\boldsymbol{\theta}}) = \mathrm{Cov}_\zeta\left[E_y(\tilde{\boldsymbol{\zeta}}_{(L)}^{\mathrm{ML}} | \boldsymbol{\zeta}_{(L)}, \mathbf{X}_{(L)};\hat{\boldsymbol{\theta}})\right] + \mathrm{E}_\zeta\left[\mathrm{Cov}_y(\tilde{\boldsymbol{\zeta}}_{(L)}^{\mathrm{ML}} | \boldsymbol{\zeta}_{(L)}, \mathbf{X}_{(L)};\hat{\boldsymbol{\theta}})\right]$$

"**线性案例**"：在此，最大似然估计量的边缘抽样协方差矩阵为

$$\mathrm{Cov}_y(\tilde{\boldsymbol{\zeta}}_{(L)}^{\mathrm{ML}} | \mathbf{X}_{(L)};\hat{\boldsymbol{\theta}}) = \hat{\boldsymbol{\Psi}} + (\hat{\boldsymbol{\Lambda}}'_{(L)}\hat{\boldsymbol{\Theta}}_{(L)}\hat{\boldsymbol{\Lambda}}_{(L)})^{-1} \tag{7.15}$$

条件抽样协方差矩阵为

$$\mathrm{Cov}_y(\tilde{\boldsymbol{\zeta}}_{(L)}^{\mathrm{ML}} | \boldsymbol{\zeta}_{(L)}, \mathbf{X}_{(L)};\hat{\boldsymbol{\theta}}) = (\hat{\boldsymbol{\Lambda}}'_{(L)}\hat{\boldsymbol{\Theta}}_{(L)}\hat{\boldsymbol{\Lambda}}_{(L)})^{-1} = \mathrm{Cov}_y(\tilde{\boldsymbol{\zeta}}_{(L)}^{\mathrm{ML}} - \boldsymbol{\zeta}_{(L)} | \mathbf{X}_{(L)};\hat{\boldsymbol{\theta}})$$

与非条件预测误差的方差相同。

对于线性随机截距模型，边缘抽样方差为

$$\mathrm{Var}_y(\tilde{\boldsymbol{\zeta}}_{(2)}^{\mathrm{ML}} | \mathbf{X}_{(2)};\hat{\boldsymbol{\theta}}) = \hat{\psi} + \hat{\theta}/n_j$$

其条件方差和预测误差的方差则为

$$\mathrm{Var}_y(\tilde{\zeta}_{(2)}^{\mathrm{ML}} | \zeta_{(2)}\mathbf{X}_{(2)};\hat{\boldsymbol{\theta}}) = \mathrm{Var}_y(\tilde{\zeta}_{(2)}^{\mathrm{ML}} - \zeta_{(2)} | \mathbf{X}_{(2)};\hat{\boldsymbol{\theta}}) = \hat{\theta}/n_j$$

## 7.6 在"线性案例"中几种记分法的关系

我们现在介绍一个非常有启发性的表达式，在"线性案例"中，该表达式连接经验贝叶斯、经验贝叶斯众数以及最大似然法。

式（7.15）中的最大似然预测量的边缘抽样协方差矩阵可写为

$$\mathrm{Cov}_y(\tilde{\boldsymbol{\zeta}}_{(L)}^{\mathrm{ML}} | \mathbf{X}_{(L)};\hat{\boldsymbol{\theta}}) = \hat{\boldsymbol{\Psi}}_{(L)} + \hat{\mathbf{A}}_{(L)}$$

其中

$$\hat{\mathbf{A}}_{(L)} \equiv (\hat{\boldsymbol{\Lambda}}'_{(L)}\hat{\boldsymbol{\Theta}}_{(L)}^{+}\hat{\boldsymbol{\Lambda}}_{(L)})^{-1}$$

$\hat{\mathbf{A}}_{(L)}$ 和 $\hat{\boldsymbol{\psi}}_{(L)}$ 分别表示对于协方差矩阵的**群组内**和**群组间**贡献。最大似然估计量的**多元信度**（multivariate reliability）就可定义为

$$\hat{\mathbf{R}}_{(L)} \equiv \text{Cov}_y(\boldsymbol{\zeta}_{(L)})[\text{Cov}_y(\tilde{\boldsymbol{\zeta}}^{\text{ML}}_{(L)})]^{-1} = \hat{\boldsymbol{\Psi}}_{(L)}(\hat{\boldsymbol{\Psi}}_{(L)} + \hat{\mathbf{A}}_{(L)})^{-1}$$

根据博科(Bock,1983)的证明,下述等式成立:

$$\tilde{\boldsymbol{\zeta}}^{\text{EB}}_{(L)} = \tilde{\boldsymbol{\zeta}}^{\text{EBM}}_{(L)} = \hat{\mathbf{R}}_{(L)}\tilde{\boldsymbol{\zeta}}^{\text{ML}}_{(L)}$$

该等式即为统计学文献中被称为**收缩**现象的多元表示(如 James and Stein,1961)。我们注意到当所估计的第一层次变量 $\hat{\boldsymbol{\Theta}}$ 大于所估计的群组间变差 $\hat{\boldsymbol{\psi}}_{(L)}$ 时,经验贝叶斯预测量被牵引向潜变量的先验期望值0。从另一个方面来说,当群组间变差大于群组内变差时(如由于群组规模较大),经验贝叶斯预测量被牵引向最大似然估计量。在极限上,即当 $\hat{R}_{(L)}=\mathbf{I}$ 时,我们得到 $\tilde{\zeta}^{\text{EB}}_{(L)}=\tilde{\zeta}^{\text{EBM}}_{(L)}=\tilde{\zeta}^{\text{ML}}_{(L)}$;这三种方法全部一致。当然,这些结果与我们选择合适的潜变量计分方法的直觉相符合。

对于随机截距模型,根据式(7.5),我们得到下述信度

$$\hat{R}_{(2)} = \frac{\hat{\psi}}{\hat{\psi} + \hat{\theta}/n_j}$$

注意,在这个三层次数值例子里,$\hat{\mathbf{R}}_{1(3)}$ 并未被定义,因为对于任何更多层次的模型来说,$\mathbf{A}_{1(3)}$ 都是奇异的。

## 7.7 专门的计分方法

在社会科学领域,经常使用假设构念(见1.3节),对此经常采用特别的计分方法,如简单地对题器或指标的数值进行加总。这些方法潜含的一个重要假设为:对得分有贡献的这些题器所测量的是同一个单维概念。对于多维概念,有时候也将这些方法应用于题器的子集上,得到"子刻度"。

### 7.7.1 原始加总分

对于连续响应、定序响应及二项响应,最常见的计分方法毫无疑问是原始加总分或未加权的加总分,其定义为

$$y_{\cdot j} = \sum_{i=1}^{n_j} y_{ij}$$

当响应为定序变量,且编码为连续整数时,所得到的得分称为 Likert 刻度(Likert,1932)。

在心理学、精神病学及相关领域中,基于问卷或结构化访谈的测量刻度通常被定义为二项或定序响应的原始加总分。研究者如果使用这种"经过校验的量表",则应根据问卷的手册来使用计分方法。但这种操作无法用于严肃的测量建模。

在某些情形下,加总分方法也存在理论上的合理性。对于连续响应变量的单维平行测量模型(见3.3.2节),所有的因子负荷相等,并且**独特因子方差**也相等(如 Jöreskog,1971b;Maxwell,1971),因此,原始加总分方法等价于得分的最大似然估计(如巴特利特法)。对于3.3.4节所讨论的二项响应变量的拉什模型,原始加总分构成估计"能力"的充分统计量。然而,单维平行测量模型和拉什模型都是非常严格的模型,在操作中不太可能成立。

一般来说,将原始加总分当做测量策略并不具有理论驱动力,并通常意味着拒绝测量建模。因而,托格森(Torgerson,1958)将加总分法当做"按照指令测量(measurement by

fiat)"的典型,并把它与根据测量建模所得到的"经典测量"相对比。在操作中,某些形式的建模是为了证明加总分的合理性。例如,通常用因子模型来研究单维性。这种用模型来为专门的计分方法提供证明的方法看上去似乎自相矛盾。

支持加总分方法的一个标准论点是,已经重复得到证明,当变量的个数较多时,加总分与采用更复杂的方法所得到的得分之间的皮尔逊相关一般来说接近于1。对于连续变量(见 Wilks,1938;Gulliksen,1950;Wang and Stanley,1970;Wainer,1976)以及二项变量(见 Muthén,1977;Kim and Rabjohn,1978),该论点都成立。这个结果具有非常光明的一面,即非常简单的方法看上去也能与用非常繁琐的方法一样有效。然而,使用高度皮尔逊相关作为证据已经遭到批评。其理由是所关注的潜在刻度的某些部分的差异,如在能力测试中作为录取的切割点,在使用相关这类描述统计量的时候被抹除了(如 Hambleton and Swaminathan,1985)。

加总分方法的一个主要限制在于,它无法直接应用到包含缺失题器的群组上。对于这样的情形,一种专门的处理方式是使用非缺失单元的群组均值对该缺失题器进行插值。最后,与基于模型的方法相比较,加总分方法无法在潜变量之间包含协变量信息或关系(回归或协方差)。

### 7.7.2 其他方法

原始加总分的另一种处理方法是从加总分中去掉那些"不显著"的题器(如见 Thurstone,1947;Gorsuch,1983)。该方法所存在的根本缺陷是对**显著**的定义非常主观。

**代表性题器**方法去掉其他所有题器,但只保留一个之前已被测量并且可能有效的题器,并将关于该题器的响应作为潜在得分(如 Rummel,1967)。该方法常用于生活质量问卷中,用一个题器去直接询问生活的质量。对这单个问题的回答经常被当做黄金标准(gold standard),并用于校验根据其他题器所推导出来的得分。很显然,代表性题器方法牺牲了很多信息,并假设数据具有极其高的质量(如 Adams,1975)。

在**因子负荷权重**方法中,将因子负荷当做权重来得到经过加权了的加总分(如 Fruchter,1954;Blalock,1960)。该方法的问题在于缺少一个理论依据。另外,和其他题器共享的影响被用于个体题器的计分,将导致冗余解(另见 Glass and Maguire,1966;Harris,1967;Halperin,1976)。

最后,使用因子计分矩阵的"线性案例"因子计分方法有时也被应用于非连续响应,这种方法也应避免使用。

## 7.8 潜变量计分和分类的使用

### 7.8.1 介 绍

关于潜变量计分以及分类存在很多应用,包括测量、能力测试、疾病制图、小域分析、医疗诊断、图像分析及模型诊断。对于连续潜变量,我们可区分两种类型的潜变量得分。我们在 7.8.2 节讨论测量的**因子得分**(factor score),在 7.8.3 节讨论**随机效应得分**(random effects score)。在 7.8.4 节,我们简略地讨论分类,在 7.8.5 节中,认为潜变量计分对于模型诊断非常有用,关于这个主题在下一章的 8.6.2 节将会再次回顾。

### 7.8.2 严格意义上的测量因子计分

关于测量的传统定义来自斯蒂文思(Stevens,1951),他将测量定义为根据某种规则对于某个属性赋值。该定义非常宽泛,在我们看来并不是很有用。许多感兴趣的现象都可以很好地被当做潜变量或因子,因而很自然地,**因子计分**(factor scoring)就受到关注。

克劳格(Clogg,1988)将测量过程定义为包含下述步骤:

(1)题器选择。

(2)尝试性的潜特质模型设定。

(3)对所保留的潜特质模型的选择。

(4)估计及解释模型参数。

(5)潜特质测量。

在计量心理学和统计学中,对于测量过程的前4个步骤已经有相当多的文献。有点奇怪甚至有些讽刺意味的是,大多数关于测量模型的研究停留在测量之前。例如,伯恩斯特(Bohrnstedt,1983)所写的名为"测量"的那一章。尽管对于测量模型给出了一个很好的介绍,但是并没有明确说明测量的本质是什么,而根据标题这正是所期望得到的。

从另一个极端来说,潜变量模型的目的有时是为了引出计分过程(如 Gorsuch,1983)。计分过程随后用于其他样本,而这就使得在随后的研究中潜变量模型变得毫无必要。然而,将计分过程应用于不同总体的整个策略存在非常大的问题。特别是,根据**因子恒等性**理论的结果(如 Skrondal and Laake,1999;Meredith,1964,1993),计分权重在不同总体中并不是恒等的。因此,我们建议根据潜变量得分所属的同一个总体来估计或"校准"模型参数,或者通过分组建模。

除了测量的需求之外,还有一些其他方面的动机需要计算因子得分。因子得分有助于模型解释。特别是当在因子模型中讨论维度性的时候,潜变量得分图通常能发挥作用(如 McDonald,1967;Etezadi-Amoli and McDonald,1983),请见图10.5和图10.7的应用。在对单元分类时因子得分也非常有用。例如,根据能力测试来录取学生以及根据精神健康测试对患者采取不同的治疗方案(如 Duncan-Jones et al.,1986;Muthén,1989a)。然而,对于这些目的而言,使用离散潜变量以及根据后验概率来对单元分类看上去更为简单。

在**适应性测试**中,因子得分也非常有用。在测试中,随着新的题器响应的获得,得分持续变化。"当前的"得分决定下一个题器的选择,从而使获得的"信息"最大化(如 Bock and Mislevy,1982)。

按照惯例,因子得分作为进一步分析的媒介。特别是在结构方程模型中,因子得分通常用于计算潜变量插值。这也是许多研究者呼吁计算因子得分的其中一个很重要的理由(如 Kim and Mueller,1978;Gorsuch,1983;Johnson and Wichern,1983)。奇姆和穆勒(Kim and Mueller,1978,p.60)认为

> 事实上,除了计量心理学以外,因子分析看上去更多的是当做创建因子刻度的方法,而非研究因子结构本身的方法。

从目前的情形来看,这个结论仍然是完全恰当的,一些软件包将因子分析放入所谓"数据削减"的菜单内,而这个菜单中的另一个选项为主成分分析。由于很多研究将因子得分当做进一步分析的媒介,因而有必要指出,这种方法是有问题的。除非在计算得分时非常小心,否则模型参数的估计将会存在偏差,且标准误会被低估(Skrondal and Laake,2001)。

估计潜变量模型的现代方法是去直接估计模型参数,而非求诸于计算出来的潜变

量。这或许可以解释关于潜变量模型计分的研究为何如此贫乏。

### 7.8.3 群组特定效应的随机效应得分

在随机效应模型中,每个群组被建构为拥有它自身的回归"参数",并且每个群组都被当做是从某个分布中抽样而得到。那么,随机效应得分就表示解释变量对群组的效应的评估。很显然,这些效应通常具有实质性的意义。在增长曲线或发展模型中,随机效应非常有用。在这些模型中,得分是对单个或一组群组画出**增长轨迹**的基础(如 Strenio et al., 1983)。图 11.1 为该模型在癫痫症发作研究的应用。

随机效应得分的另一个应用为群组的分类与排序。这方面的例子包括将不同组织根据其随机效应得分分成高效率和低效率组织(Aitkin and Longford,1986),根据收入的性别差距将不同行业进行排序(Kreft and de Leeuw,1994),以及根据成绩对学校进行排序(Goldstein and Spiegelhalter,1996)等。为了比较两个单元的随机效应,适合采用的标准误为预测误差的标准差。然而,排序问题则具有非常大的差异性,其精确性无法简单地用标准误来表示(Goldstein and Spiegelhalter,1996)。

在元分析中,随机效应模型也常用于将不同研究中的效应估计值进行合并。在这种情形中,很自然地会采用将原来的效应估计值替换为随机效应模型的后验均值的方法,以从其他研究中"**借力**"。通常使用后验标准差来表示置信区间,而在连续变量的情形中,置信区间表示预测误差的方差。在 9.5 节的元分析中,我们使用该方法研究尼古丁口香糖对戒烟的效果。

在疾病制图和小域估计中,经常使用随机效应计分。对于总体规模小的区域来说,对发生频率的估计非常不准确,所得到的图形具有"噪点"。从其他区域借力可获得更可靠、更平滑的图形。理想来说,在这种情形中,模型应该利用空间信息;请见 Clayton and Kaldor(1987)、Langford et al.(1999)及 11.4 节。

### 7.8.4 分 类

在某些情形中,潜变量的本质是离散的,典型的例子如在医学诊疗中,患者是否具有某种疾病。在这种情形下,为了开出正确的处方,分类是至关紧要的。在市场营销中,有时按照某组属性,将顾客划分为是否属于某个"细分市场",以进行有目的的广告投放(如 Wedel and Kamakura,2000)。

在大多数情况下,潜变量被当做连续的,但是对于决策制定来说,分类则是必需的。出于此目的,就需要设定一个离散潜变量模型(如潜类模型)。我们可使用经验贝叶斯众数来对单元进行分类,而不必将一个主观的阈值设定在连续得分上。在教育中,当关注的主题是一个科目是否被掌握,而非在一个连续刻度上的能力时(如 Bergan,1988; MacReady and Dayton,1992),这种方法也被采用。

在图像分析中,一个常见问题是图像分割或修补,将像素(处于方形栅格上的图像元素)或三维像素(处于立方体上的体积元素)分成不同区域。例如,在脑图像中描绘出脑肿瘤。对潜在区域设置标签或潜类,有时可设定一个空间模型,如马尔可夫随机场(如 Besag,1986),其中响应满足正态分布并且条件依赖于潜类,表示"噪声"。图像分割就可通过诸如"模拟退火"(如 Geman and Geman,1984)这样的算法来找到模式后验区域标签来实现。

在本书中,我们主要讨论的潜类模型是在任意单元的真实分类未知的条件下的分类。如果单元的部分子样本的真实分类已知(有时称为"训练集"或"校验样本"),潜类

模型可扩展成如 14.3 节所示的模型。维德尔(Wedel,2002a)讨论了一种情形,其中对于训练集来说,其“核心变量”(响应)与“伴随变量”(解释变量)已被观测到,但对于需要分类的新单元,只有伴随变量被观测到。另一个情形是对单个的定类响应变量 $y$ 进行预测,而离散潜变量被当做神经网络中的干预变量的“隐层(hidden layer)”(请见 Vermunt and Magidson(2003b)对此类模型的综述)。

与潜变量计分一样,潜分类有时候也被当做“观测到的”变量用于后续分析中。维德尔(Wedel,2002b)所提出的一个类似的方法是用估计的后验概率对解释变量进行回归。这两种方法都存在问题,并应予以避免(如 Croon,2002)。至于对连续潜变量来说,直接对潜变量与观测响应及(或)协变量之间的关系进行建模更可取。在 14.3 节,我们估计了一个包含协变量的潜类模型,在 13.5 节将潜类对协变量进行回归。请注意,在这个例子中,对潜类的估计使用了包含在模型中的全部变量的信息,而不仅是“测量”潜类的那部分题器的信息。

### 7.8.5 模型诊断

潜计分的一个重要应用为模型**诊断**。像研究观测变量那样,用潜计分来研究统计模型背后的假设是一件诱惑人的事情。然而,请记住,由于所预测的得分的分布与理论上的潜变量的分布并不相同,因此使得这种方法存在问题,特别是对非正态响应来说。另请见 8.6.2 节。

在异常值探测中,潜得分可被当做估计的残差,如比较得分及其近似抽样标准差。关于潜计分在一般潜变量模型诊断中的使用将 8.6.2 节中讨论;另请见 11.3.3 节的应用。

## 7.9 总结与进一步阅读

对于在本章所讨论的所有基于计分和分类的模型,只要响应缺失可被假设为随机性缺失(MAR),缺失不会带来任何问题。

到目前为止,对于连续潜变量赋值最常用的方法为经验贝叶斯预测。该方法的优点在于能使得损失函数均方差最小化(如果模型参数被假设为已知),并且是非条件无偏。当在某些单元上存在稀疏信息时,条件有偏或与此方法相关联的收缩可被看做是一个优点。

有些情况下也可采用最大似然法,因为相比起经验贝叶斯,其得分为条件无偏。然而,该方法与模型假设并不一致,对于信息不充分的群组无法得到预测。甚至,该方法无法用于真正的多层次模型。

关于标准误的定义存在多种方式。其中,最常用的为后验标准差,在连续变量的情形中,它等于预测误差的标准差,以及边缘抽样标准差。其要点在于,无论采用哪种计分方法,必须使用与其相适合的标准误。

对于离散潜变量,最常使用的分类方法为经验贝叶斯众数,因为它可以使期望分类错误率最小化。

与潜变量模型的估计相反,关于潜计分和分类的文献相对缺乏。关于经验贝叶斯,比较有用的书为 Maritz and Lwin(1989)和 Carlin and Louis(1998)。斯特尼奥等(Strenio et al.,1983)针对线性随机系数模型的经验贝叶斯预测做了一个综述。关于巴特利特法以及因子分析的回归方法的评述请见 Lawley and Maxwell(1971,第 8 章)。

附录 软 件

我们列出部分能够对潜变量赋值的软件，它们使用不同的方法。我们不提供这些软件的地址或链接，因为这些信息更新得很快，也很容易从互联网获得。

• 经验贝叶斯

——aML 处理多层次模型以及多过程模型，使用积分法(Lillard and Panis,2000)。

——BILOG-MG 处理二进制对数概率题器响应模型，使用积分法(Zimowski et al., 1996;Du Toit,2003)。

——gllapred,gllamm 的估计命令，处理广义线性潜在及混合模型，使用自适应积分法(Rabe-Hesketh et al.,2001b,2004c)。

——MIXNO,MIXOR,MIXREG,MIXPREG 和 MIXSUR 处理双层次广义线性混合模型，使用积分法(Hedeker and Gibbons,1996ab;Hedeker,1999)。

——TESTFACT 处理多维度正态概率因子模型，使用自适应积分法(Bock et al., 1999)。

• 经验贝叶斯众数

——计分：

* BILOG-MG 处理二进制对数概率题器响应模型(Zimowski et al., 1996; Du Toit, 2003)。

* HLM 处理多层次广义线性混合模型，使用 PQL 或 LaPlace6(Raudenbush et al., 2001)。

* MLwiN 处理广义线性混合模型，使用 PQL(Rasbash et al.,2000)。

* Mplus 处理包含连续、二项、定序以及删截响应的结构方程模型(Muthén and Muthén,1998,2003)。

* SAS NLMIXED 处理双层次广义线性混合模型(Wolfinger,1999)。

——分类：

* Mplus 处理包含连续、二项、定序以及删截响应的潜类模型(Muthén and Muthén, 1998,2003)。

* Latent GOLD 处理大多数响应类型(Vermunt and Magidson,2000,2003a)。

* gllamm(后验概率)处理 Stata 中的广义线性潜在及混合模型(Rabe-Hesketh et al., 2001b,2004c)。

• 最大似然估计

——BILOG-MG 处理二进制对数概率题器响应模型(如 Du Toit,2003)。

# 8 模型设定与推论

## 8.1 介　绍

在第六章，我们已经讨论了对统计模型进行参数估计的问题，但没有考虑为什么设定这个模型。本章我们讨论如何寻找一个合适的模型，而这可能是更为困难的任务。

在开始讨论模型设定之前，我们在8.2节对统计建模进行回顾。特别是，我们讨论不同类型的统计模型的角色与目的，以及其建模策略。在8.3节，我们对最大似然推论进行回顾，最大似然推论在评估估计的不确定性方面非常有用，并且构成了在8.4节所讨论的模型选择的“相对拟合”标准的基础。在完全依靠相对拟合方面还存在争论，我们所能下的结论为，一个模型虽然看似优于其他竞争模型，但是在绝对的尺度上，一个模型好多少或差多少却很难说。此外，对备择模型的错误设定可能会造成模型选择无效。因此，在8.5节我们采用“**全局绝对拟合**(global absolute fit)”标准来作为分析最优模型好到什么程度的方法。与全局绝对拟合标准相反，“**局部绝对拟合**(local absolute fit)”标准不仅可用于发现一个模型是否不适当，还可用于诊断模型在何处设定错误。诊断方法将在8.6节讨论。

本章节的安排似乎意味着建模过程按照下述顺序排列：①从多个模型中选择“最优”模型；②评估所选模型(如果可行的话)；③用诊断法找到设定错误的地方，而这可能意味着我们需要采用别的方式来阐述模型并重新回到步骤①。但是，这些顺序并没有任何理论依据，其他顺序也是有可能的。

## 8.2 统计建模

### 8.2.1 统计模型类型

广泛地讲，在经验研究的**数学模型**应用中，变量之间的关系被形式化为一个或多个确定性方程。在由确定性法则决定的关系中，这些模型是合适的，如同在自然科学中的那样，典型的例子比如牛顿的运动定律。与此相反，**统计模型**作为数学模型不仅包括确定性部分，还包括随机部分。这个随机部分可能表示测量误差，使得在研究确定性规律的时候统计模型也仍然有用。更重要的是，它也可表示“自然变异”或随机因果律，如孟德尔遗传。最后，随机部分也可能反映我们对于所研究的经验现象背后的确定性规律的

知识不够完备。需要注意的是,一些统计工具与统计模型无关,因为它们没有包含随机部分。例如,主成分分析仅仅是对数据的正交转换。

在统计学文献中区分了两种类型的统计模型。我们采用考克斯(Cox, 1990)所使用的术语:**实质模型**和**经验模型**。勒曼恩(Lehmann,1990)介绍了奈曼(如 Neyman, 1939)所提出的类似的区分,即奈曼区分成"解释模型"与"**插值公式**(interpolatory formulae)",巴克斯等(如 Box et al.,1978)则区分成"理论模型"或"机制模型"与"经验模型"。

**实质模型**

最有吸引力的统计模型莫过于直接与所关注的主题及背景信息相关联的实质模型,这种模型能获得对事物的理解与解释,也就是说,是回答哲学上的"为什么"。通常来说,研究者相信存在一个"真实的"模型,是这个"真实"模型生成了数据。

"**直接实质模型**(Directly substantive models)"根据外显的机制来解释观测到的是什么,通常通过未被直接观测到的量以及一些理论概念,如所研究的系统如何"运行"。在自然科学中,可能会存在一个或多数不多的几个竞争理论试图去解释观测到的现象,一般使用带有特定函数形式的定理性的关系。奈曼最喜欢使用的例子为孟德尔遗传规律。为了"检验"一个理论,研究者有时在可控制的条件下采用随机性分配进行所关注的"试验"。

有一种较弱类型的实质模型(如图像或结构方程模型)仅仅是关于独立性的实质性假设,即某些变量可被设定为在给定其他变量条件下的条件独立。例如,在 1.3 节中所提的结构方程模型假设"**纯粹的中介作用**",即其中一个变量对结果变量仅有通过中介变量的非直接效应,而无直接效应。这些模型在社会科学和医学中非常有用,在这些学科中,经常会有多个不太精确的甚至相互矛盾的关于模糊现象之间关系的想法。通常信息来自于过往对此问题的经验研究以及相关问题。然而,在基于观测设计的研究中,研究者仅仅是被动的经验过程的观测者,无法对因果性做进一步探究。稍微强一点的设计为准实验法(如 Cook and Campbell,1979),研究者可以变更试验,但由于某些原因无法进行随机化。

由于实质性统计模型是关于数据生成过程的简化表示,因此,根据模型来直接生成模拟数据应该是可行的。一个好的实质模型应该是简洁的,并且能够抓住数据生成过程的主要特征;它既不能过于复杂也不应过于简单。过于复杂的模型毫无用处,因为它仅仅是实际数据的镜像,而非潜在的过程,并且很可能无法表征同一过程的其他实现。而从另一方面来说,过于简单的模型无法抓住数据生成过程的主要特性,从而导致错误的推论。

考克斯和魏玛斯(Cox and Wermuth,1996)认为一个令人满意的实质性统计模型应该:

(1)与背景知识建立连接。

(2)与以往的研究相关联。

(3)能够阐明数据生成过程。

(4)主要参数是关于主题的清晰的实质性解释。

(5)合理地设置偶然性,以便进行关于精度的有意义的评估。

(6)有足够的拟合度。

**经验模型**

在很多应用研究中,背景信息相对缺乏,较常使用的是经验模型。根据巴克斯(Box)的定义,经验模型被当做行动的向导,侧重于**预测**。模型更多的是对特定情景提供指引,

利用一切细节和背景,使得好的近似结果仅可能出现在所关注的领域。在一大类出于便利而选择的模型中,无须对潜在情景的假设,即可完全基于数据而得到经验模型。研究者不用去假设存在真实模型,而是从一系列可能有用的模型中去寻找模型。它的一个现代版本为布雷曼(如 Breiman,2001)所提倡的**算法模型**(algorithmic model),这个模型是一种黑箱策略,即关注于根据输入来对输出进行预测的算法的有效性和灵活性。模型并不关注黑箱本身,并把黑箱的内容当做不可知。典型的例子如神经网络或"**回归森林**(regression forests)"。

考克斯(Cox,1990)提出了一种不太极端的经验模型,该模型并不是基于特定的实质性考虑,其目的是为了表示被认为存在的理想形式的依赖,通常为"平滑"依赖。根据考克斯,经验模型的首要并且最常见的角色为估计效应及其精度。回归模型的普遍使用就是一个典型例子,如在流行病学中根据对数概率回归来估计关联度。必须指出的是,这些类型的经验模型并不是完全没有实质性考虑,如在流行病学中对混淆因子的选择。经验模型的另一个角色为"修正数据的不足",如测量误差、数据缺失及复杂抽样。这种不太极端的经验模型概念与较弱类型的实质模型的想法非常相似。

## 8.2.2 建模策略

**模型设定**的问题,即设定一个低维度的参数统计模型,是费舍尔(Fisher,1922)所提到的统计所需要面对的第一个问题。有意思的是,他关于设定的讨论仅仅为一个段落的第一句:

至于设定的问题,对于实践统计学家来说这是至关重要的……

勒曼恩(Lehmann,1990)将费舍尔的观点理解为可以没有关于建模的理论,也可以没有建模的策略,但是对于每个问题都必须从其本质上给予考虑。

然而,在实际操作中,通常采用我们前面所列的 4 种建模策略的其中一种(见 Jöreskog(1993)所作的类似的区分)。对于实质模型来说,一个很自然的建模策略为包含一个或两个模型的**完全验证**(strictly confirmatory)方法。例如,在测量建模中,研究者可能希望确定某个特别的独立群组结构(见 3.3.3 节)是否成立,他要根据某个绝对的拟合标准来判断是否保留或拒绝这个模型,请见 8.5 节。临床试验经常包含两个模型,其中零模型通常为药物不存在效应,而备择模型则是该药物存在效应。模型选择通过假设检验来进行。重要的是,这些模型是在事先被设定的(在 ANOVA 中被称为"设计对照"),而不是根据检验所用的数据而提出。然而,由数据得来的模型或"理论"也非常常见,并且为了使得结论更加可靠,仿佛它们是事先就设定好了的。在药物开发中,为了避免这样的不当(可能会导致致命的结果),在开始临床试验前就准备一份详细的分析计划并严格按此计划进行检验成为普遍的做法。

另一种对实质模型建模的策略为**竞争模型**(competing models)法,即设定一定数量的备择模型,并从中选择。当对于需要解释的现象存在多个竞争的理论,且研究者需要对不完善的理论进行批判时,这是比较合适的。

对于经验模型,或较弱的实质模型,一个自然的建模策略为**模型生成**(model generating)法,即基于已有的背景信息来设定一个初始的试验性模型。如果根据诊断与/或拟合标准,该模型不能被接受,则根据背景理论或者出于对数据的更佳拟合的目的来修改模型。这一系列过程包括模型设定、估计、与数据相验证以及再设定,不断迭代,直到模型可以被接受。

对于经验模型，典型的建模方法为**完全探索法**(strictly exploratory)，即根据数据来"推导"模型。在操作上，它需要从大量的竞争模型开始，使用纯粹的统计标准从中进行选择。一个普遍的例子如最优子集线性回归，这个模型协变量的个数很多，需从协变量的子集中进行选择。对于预测来说，如果与某些形式的交叉验证(进一步讨论见后)相结合以避免过度拟合的话，这种方法可能会得到有用的模型。

重要的是，该探索性方法不应用于推导实质模型。众所周知，这种来自"数据挖掘"的结果最多仅具有启发性，并且应当用验证性的方法用独立数据进行评估。弗里德曼(Freedman，1983)证明，理论也可推导自纯粹的噪声数据中。他模拟了100个单元的51个独立标准正态变量，并将最后一个变量当做响应变量，其他剩余变量当做线性回归中潜在的协变量。仅仅选择那些在25%的水平上系数显著的协变量，也能得到与那些在5%的水平上系数显著的模型一样"令人信服"的结果。艾霍恩(Einhorn，1969)针对"自动化的交互作用探测"、回归分析、因子分析及非度量型的多维尺度做了类似的批判。他的论文"行为科学中的炼金术"的结论非常值得摘录一下：

> 必须清楚的是，不是依据理论而是根据强大的数据分析技术，将导致大量的一类错误(Type I error)。正如古代的炼金术士无法将金属变成金子一样，现代的研究者也无法依赖于计算机来使他的数据变成有意义和有价值的科学信息。

尽管验证性方法与探索性方法的区别及其蕴涵适用于任何类型的建模，我们仍然从因子分析的角度来讨论一下。

例：验证性因子分析与探索性因子分析

验证性因子分析(CFA)是一种假设的分析过程，它被设计为用于检验题器和因子之间关系的假设，其中因子的个数和解释是事先给定的。因此，在验证性方法中，特定的参数被设定为规定的数值。

而探索性因子分析(EFA)则作为一种归纳式的方法，用于发现可解释题器之间变差的最优的因子集合(见 Mulaik，1988b；Holzinger and Harman，1941)。穆拉克(Mulaik，1988ab)指出在归纳法范畴中 EFA 无法实现其目标，共有3个理由：首先，在不做先验假设的情况下，并不存在根据经验来获取知识的理性最优的方式。其次，由于因子不确定性(Guttmann，1955)，对一个探索性因子分析模型的解释不存在唯一性。最后，当原理上无法证伪一个模型的时候，也难以证明该模型的结果。麦当劳(McDonald，1985：102)指出了 EFA 的模糊性：

> 在探索性方法中，或许可以认为，我们的做法并不一致。由于结构简单，我们首先拟合了一个有很多参数并且没有约束的模型。然后，我们将此结果转换为近似于简单结构的甚至可能更拙劣的等价拟合，并认为现在有了一个参数更少的模型。然而，在此简单结构中的参数个数可能与总体中确切的零相一致，也可能不一致。如果相一致，我们只能估计非零值。如果不一致，那么实际上，我们根本没有简单结构。

也有学者认为 EFA 的结果可能具有启发性和启示性的价值(如 Anderson，1963)，并可能发现一些假设，从而可用其他多元分析方法(Hotelling，1957)以及新的数据进行更有目的的检验。然而，试图根据 EFA 来获得有意义的假设这个前景非常黯淡，正如阿姆斯特朗(Armstrong，1967)(另见 Mukherjee，1973)所证明的。阿姆斯特朗认为有意义的 EFA 仅当存在足够的先验信息时才有可能，而在这个条件下，应首先采用 CFA(另见1.3节)。

在量表设计中使用验证性模型也可能受到批评。有些情况下，研究者会偏好某些模

型,如3.3.4节所讨论的单维拉什模型,并且为了确保得到好的拟合,抛弃与模型相矛盾的一些题器。戈德斯坦恩(Goldstein,1994)指出,这种方法是有问题的,因为好的拟合是对潜变量单维性的有力支持。

从现在开始,我们假设我们的目的是采用一系列建模策略(包括从完全验证法到模型生成法)来选择一个较弱的实质模型。值得一提的是,在开始形式化的统计建模前,通常可基于或多或少的一些启发制订一些重要的方案,如:

- 选择一个模型类型(如多层次模型)
- 变量的"因果"排序:

——回归模型:通常分为一系列解释变量与一系列响应变量(有时候只有一个)。

——结构方程模型:对解释变量、中介变量和响应变量的更细致的安排。

——潜变量模型:选择"测量"潜变量的变量。

- 设定概率分布

根据这些方案,统计建模的任务通常是帮助确定**模型形式**,如:包含哪些解释变量、交互作用,在结构方程模型中包含哪些"路径",以及哪些题器测量哪些潜变量。

所采用的建模策略通常取决于模型类型以及研究对象所属的领域。例如,对于生物统计中的线性混合模型,沃贝克和莫莱恩伯(Verbeke and Molenberghs,2000,第9章)提出下述建立模型的顺序。首先,对于固定部分寻找一个预备模型,选择协变量以及决定它们与响应的函数关系。其次,对于随机部分寻找一个预备模型,并决定哪些效应被当做随机的。最后,对于第一层次的误差,寻找一个合理的模型,如是否应该设定异方差性与/或自相关。

## 8.3 推论(基于似然)

### 8.3.1 基本参数估计和结构参数估计的属性

在用最大似然估计参数之后,接下来的问题就为参数估计的属性。首先考虑所估计的基本参数 $\hat{\boldsymbol{\vartheta}}$。由于 $\hat{\boldsymbol{\vartheta}}$ 为ML估计量,根据考克斯和亨克莱(Cox and Hinkley,1974),在适当的正则条件下,它具有很多很好的理论属性。例如,$\hat{\boldsymbol{\vartheta}}$ 为一致的**渐进正态**,并且是**渐进有效**。接下来,考虑结构参数 $\boldsymbol{\theta}=\mathbf{h}(\boldsymbol{\vartheta})$ 的估计量 $\hat{\boldsymbol{\vartheta}}$。首先,众所周知,由于ML估计量在转换中是守恒的(如Cox and Hinkley,1974:287),因此 $\boldsymbol{\theta}$ 的ML估计量,$\hat{\boldsymbol{\theta}}$ 由下式给出

$$\hat{\boldsymbol{\theta}} = \mathbf{h}\hat{\boldsymbol{\vartheta}} \tag{8.1}$$

由此得出,$\hat{\boldsymbol{\theta}}$ 继承了 $\hat{\boldsymbol{\vartheta}}$ 的渐进最优属性。

鲁宾(Rubin,1976)证明,当响应是随机性缺失(MAR)时,最大似然估计量可获得一致性。这要求一个响应缺失的概率并不依赖于该响应本应被观测到的数值,尽管响应可能依赖于模型中所包含的协变量或其他响应。重要的是,并没有要求响应为完全随机性缺失(MCAR),即要求响应不依赖于协变量、观测到的响应或缺失响应。雷托(Little,1995)指出,当缺失依赖于协变量时,一些方法仍然有效,如GEE法,尽管通常它们被认为要求满足MCAR假设。

在纵贯性数据的随机效应模型中,雷托(Little,1995)区分了依赖于协变量的退出(covariate dependent dropout)、随机性退出(missing at random dropout)、不可忽略的基于结

果的退出（nonignorable outcome-based dropout）及不可忽略的基于随机系数的退出（nonignorable random-coefficient-based dropout）。设 $\mathbf{y}_j$ 表示观测到的和未观测到的（缺失）响应 $\mathbf{y}_j' = [\mathbf{y}_{\text{obs},j}, \mathbf{y}_{\text{mis},j}]$ 的向量，$\mathbf{r}_j$ 为在单元 $j$ 上缺失指标的向量。这几种退出类型可被定义为：

- 依赖于协变量的退出

$$\Pr(\mathbf{r}_j \mid \mathbf{y}_j, \mathbf{X}_j, \boldsymbol{\zeta}_j) = \Pr(\mathbf{r}_j \mid \mathbf{X}_j)$$

- 随机性退出

$$\Pr(\mathbf{r}_j \mid \mathbf{y}_j, \mathbf{X}_j, \boldsymbol{\zeta}_j) = \Pr(\mathbf{r}_j \mid \mathbf{y}_{\text{obs},j}, \mathbf{X}_j)$$

- 不可忽略的基于结果的退出

$$\Pr(\mathbf{r}_j \mid \mathbf{y}_j, \mathbf{X}_j, \boldsymbol{\zeta}_j) = \Pr(\mathbf{r}_j \mid \mathbf{y}_{\text{obs},j}, \mathbf{y}_{\text{mis},j}, \mathbf{X}_j)$$

- 不可忽略的基于随机系数的退出

$$\Pr(\mathbf{r}_j \mid \mathbf{y}_j, \mathbf{X}_j, \boldsymbol{\zeta}_j) = \Pr(\mathbf{r}_j \mid \mathbf{y}_{\text{obs},j}, \mathbf{X}_j, \boldsymbol{\zeta}_j)$$

我们认为这个分类方式对于包含非单调缺失数据模式的一般潜变量模型也是有用的（在纵贯性数据中，表现为断续性的缺失）。

如果缺失依赖于协变量或随机（并且实质性过程的参数与缺失过程的参数有区别），推论可仅仅基于观测到的响应的似然（实质性过程）。这是因为实质性过程和缺失过程的联合似然可以被分解为单独的部分。然而，对于后面那两种不可忽略的缺失过程来说，这个结果并不成立，必须对实质性过程和缺失过程联合建模（如 Heckman, 1979; Hausman and Wise, 1979; Wu and Carroll, 1988; Diggle and Kenward, 1994）。更详尽的讨论请参考 Little and Rubin (2002)。

### 8.3.2 基于模型的标准误

$\hat{\boldsymbol{\vartheta}}$ 的渐进协方差矩阵为

$$\text{Cov}(\hat{\boldsymbol{\vartheta}}) = -\text{E}(\mathbf{H}(\hat{\boldsymbol{\vartheta}}))^{-1} \cong -\mathbf{H}(\hat{\boldsymbol{\vartheta}})^{-1} \qquad (8.2)$$

其中，$-\text{E}(\mathbf{H}(\hat{\boldsymbol{\vartheta}}))$ 为费舍尔**信息**或期望信息，而 $-\mathbf{H}(\hat{\boldsymbol{\vartheta}})$，即对数似然的负海森矩阵，为**观测到的信息**。由于大数定理，观测到的信息近似于期望信息。

出于以下 3 个方面的原因，我们使用观测到的信息，而非期望信息：第一，纯粹是操作层面的。观测到的信息是牛顿-拉夫逊算法的副产品，因而当在参数估计中用该算法时，这是自然而然的选择。第二，正如莱尔德（Laird, 1988）和希拉奇特（Schluchter, 1988）以及其他人所指出的，当缺失数据满足 MAR 但不满足 MCAR 时，使用期望信息将存在问题。因为那样的话，为了获得正确的期望信息，就必须结合数据缺失的机制。第三，艾弗朗和亨克莱（Efron and Hinkley, 1978: 459）认为观测到的信息比期望信息"更接近于数据"（另见 Kendall and Stuart, 1979），并且与贝叶斯分析更加一致。

将多元德尔塔法（如 Serfling, 1980）应用于式（8.1）的函数，我们可根据所估计的基本参数推导出所估计的结构参数的协方差矩阵为

$$\text{Cov}(\hat{\boldsymbol{\theta}}) \cong \left(\frac{\partial \mathbf{h}(\boldsymbol{\vartheta})}{\partial \hat{\boldsymbol{\vartheta}}}\right) \text{Cov}(\hat{\boldsymbol{\vartheta}}) \left(\frac{\partial \mathbf{h}(\boldsymbol{\vartheta})}{\partial \hat{\boldsymbol{\vartheta}}}\right)'$$

### 8.3.3 稳健标准误

总的对数似然为处于顶层的群组贡献的总和

$$l(\boldsymbol{\vartheta}) = \sum_{z=1}^{n^L} l_z(\boldsymbol{\vartheta})$$

其中

$$l_z(\boldsymbol{\vartheta}) = \ln g^{(L)}(\mathbf{y}_{z(L)} \mid \mathbf{X}_{z(L)};\boldsymbol{\vartheta})$$

因此最大似然估计量满足下述似然方程

$$\mathbf{g}(\hat{\boldsymbol{\vartheta}}) = \sum_{z=1}^{n^{(L)}} \mathbf{g}^z = \mathbf{0} \tag{8.3}$$

其中,$\mathbf{g}^z$ 为 $L$ 层次上第 $z$ 个单元的得分向量,即

$$\mathbf{g}^z = \frac{\partial l_z(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}}$$

使用德尔塔法,我们可以将 $\mathbf{g}(\hat{\boldsymbol{\vartheta}})$ 的协方差矩阵写为

$$\mathrm{Cov}[\mathbf{g}(\hat{\boldsymbol{\vartheta}})] = \left(\frac{\partial \mathbf{g}(\hat{\boldsymbol{\vartheta}})}{\partial \boldsymbol{\vartheta}}\right)\mathrm{Cov}(\hat{\boldsymbol{\vartheta}})\left(\frac{\partial \mathbf{g}(\hat{\boldsymbol{\vartheta}})}{\partial \boldsymbol{\vartheta}}\right)'$$

对 $\mathrm{Cov}(\hat{\boldsymbol{\vartheta}})$ 求解,得

$$\begin{aligned}\mathrm{Cov}(\hat{\boldsymbol{\vartheta}}) &= \left\{\frac{\partial \mathbf{g}(\hat{\boldsymbol{\vartheta}})}{\partial \boldsymbol{\vartheta}}\right\}^{-1}\mathrm{Cov}[\mathbf{g}(\hat{\boldsymbol{\vartheta}})]\left\{\left(\frac{\partial \mathbf{g}(\hat{\boldsymbol{\vartheta}})}{\partial \boldsymbol{\vartheta}}\right)'\right\}^{-1} \\ &= \mathbf{H}^{-1}\mathrm{Cov}[\mathbf{g}(\hat{\boldsymbol{\vartheta}})]\mathbf{H}^{-1}\end{aligned} \tag{8.4}$$

其中,$\mathbf{H} \equiv \mathbf{H}(\hat{\boldsymbol{\vartheta}})$ 为参数估计的对数似然海森矩阵。如果该模型正确,那么 $\mathrm{Cov}[\mathbf{g}(\hat{\boldsymbol{\vartheta}})] = -\mathrm{E}(\mathbf{H}) \cong -\mathbf{H}$,使得 $\mathrm{Cov}(\hat{\boldsymbol{\vartheta}}) = -\mathrm{E}(\mathrm{H})^{-1} \cong -\mathbf{H}^{-1}$,如式(8.2)所示。

除了依赖于正确设定的模型,我们还可利用式(8.3)中 $\mathbf{g}(\hat{\boldsymbol{\vartheta}})$ 为均值为 $\mathbf{0}$ 的独立得分向量 $\mathbf{g}^z$ 的总和,使得经验协方差矩阵变为

$$\mathrm{Cov}[\mathbf{g}(\hat{\boldsymbol{\vartheta}})] = \frac{n^{(L)}}{n^{(L)}-1}\sum_{z=1}^{n^{(L)}} \mathbf{g}^z(\mathbf{g}^z)' \tag{8.5}$$

将式(8.5)代入式(8.4)得到所谓的夹挤估计量;例如见哈伯尔(Huber,1967)与怀特(White,1982)。这可被当做估计值的基于设计的抽样分布的协方差矩阵估计量,而估计值被定义为数值的(隐)函数(Binder,1983)。夹挤估计量在广义估计公式中非常流行(见 6.9 节),对于包含抽样加权的复杂抽样数据,其对数似然贡献采用逆选择概率[1]进行加权,得到"伪似然"。

如果多层次模型中的顶层单元不是相互独立,而是属于相互排斥的群组 $n_c$,且其标记集合为 $C_m, m=1,\cdots,n_c$,那么

$$\mathrm{Cov}[\mathbf{g}(\hat{\boldsymbol{\vartheta}})] = \frac{n_c}{n_c-1}\sum_{m=1}^{n_c}\left(\sum_{z\in C_m}\mathbf{g}^z\right)\left(\sum_{z\in C_m}\mathbf{g}^z\right)' \tag{8.6}$$

证明请见伍德里奇(Wooldridge,2002,13.8.2 节)与威廉姆斯(Williams,2000)。穆森和萨托拉(Muthén and Satorra,1995)提出在逆概率加权的群组数据的结构方程模型中采用此方法。

除了夹挤估计量以外,还可采用再抽样法,如**自助法**(bootstrap)和**刀切法**(jackknife)。迈尔等(Meijer et al.,1995)和巴辛等(Busing et al.,1994)讨论了线性双层次模型的参数和非参数自助法。存在两种类型的非参数自助法:一种基于对案例再抽样,

---

1 选择概率的倒数。为了下文方便,使用"逆选择概率"。

另一种基于对残差或误差再抽样。对多层次数据来说,这两种方法都不够直接。对于"案例自助法"来说,是对群组以及群组中的单元再抽样,还是对群组再抽样,又或者是仅仅对群组中的单元再抽样,这并不完全清楚。至于"误差自助法",由于收缩(另见 Carpenter et al., 1999)的缘故,如何估计处于高层次的残差也并不清楚。帕特森等(Patterson et al.,2002)将刀切法应用于包含样本权重的潜类模型,巴辛等(Busing et al., 1994)则将其应用于双层次线性模型。

## 8.3.4 似然比值、瓦尔德(Wald)检验及得分检验

设 $M_1$ 和 $M_2$ 表示两个竞争模型,其基本参数个数分别为 $v_1$ 和 $v_2$,并用 $\boldsymbol{\vartheta}_{M_1}$ 和 $\boldsymbol{\vartheta}_{M_2}$ 分别表示基本参数。假设 $M_2$ 嵌套于 $M_1$ 中,因此,约束添加在 $M_1$ 的结构参数上,从而得到模型 $M_2$,且其基本参数个数少 $v_1 - v_2$ 个。设两个模型对数似然最大值分别用 $l(\hat{\boldsymbol{\vartheta}}_{M_1} | \mathbf{y}, \mathbf{X})$ 和 $l(\hat{\boldsymbol{\vartheta}}_{\mathrm{M}_2} | \mathbf{y}, \mathbf{X})$ 表示。

传统的似然比值检验可用下述统计量得

$$\mathrm{D}_{M_1M_2} = 2[l(\hat{\boldsymbol{\vartheta}}_{M_1} | y, \mathbf{X}) - l(\hat{\boldsymbol{\vartheta}}_{M_2} | y, \mathbf{X})] \quad (8.7)$$

根据正则条件,它为渐进的 $\chi^2$ 分布,自由度为 $v_1 - v_2$(如 Cox and Hinkley, 1974)。

根据 $\mathrm{Cov}(\hat{\boldsymbol{\vartheta}}_{M_1})$,可推导出瓦尔德检验,根据 $\mathrm{Cov}(\hat{\boldsymbol{\vartheta}}_{M_2})$ 可推导出拉格朗日乘数或得分检验。这些检验统计量,在对单个模型(如 $M_1$ 或 $M_2$)进行估计时是必需的,可被当做是似然比值检验统计量的二次近似。众所周知,瓦尔德检验和拉格朗日乘数为似然比值检验的渐进等价(如 Cox and Hinkley, 1974; Buse, 1982; Engle, 1984)。

然而,在有限样本的情形当中,对检验统计量的选择非常重要。如果用估计参数临域内的一个二次函数无法很好地近似于对数似然,那么瓦尔德检验效果就会很差。豪克和多纳(Hauck and Donner, 1977)证明在对数概率回归中可能会发生这样的情况。请注意,与似然比值检验不同,对参数进行非线性转换时,瓦尔德检验并不守恒,有些转换优于其他转换。如果瓦尔德检验与似然比值检验产生不同的结果,一般倾向于选择似然比值检验。

根据中心极限定理,得分检验也是可行的,而不仅仅是作为似然比值检验的近似(如 Pawitan, 2001:235)。在某些情形中,得分检验优于似然比值检验。得分检验优于瓦尔德检验的一个方面为它关于参数的非线性转换守恒。我们建议参考帕维坦(Pawitan, 2001)对于似然理论所做的非常好的并且容易理解的解释。

然而,如果零假设是关于参数空间的边界的话,这个零假设正是违背正则条件的,那么似然比值、瓦尔德及得分检验统计量的标准渐进结果并不成立。常见的例子比如检验一个或多个方差成分是否为 0 的零假设。请考虑下述两个双层次模型 $M_1$ 和 $M_2$,其中 $M_1$ 多出一个方差 $\psi_{kk}$ 以及 $M$ 个协方差 $\psi_{mj}, j = 1, \cdots, M, j \neq k$。由于 $\psi_{kk}$ 为非负数,在 $M_1$ 中,它处在参数空间的边界内,$\psi_{kk} = 0$。正确的似然比值统计量的渐进分布为处于 0 点的 50:50 的点质量混合,其 $\chi^2$ 分布为 $M + 1$ 个自由度(Moran, 1971; Miller, 1977; Self and Liang, 1987; Berkhof and Snijders, 2001)。因此,通过计算对应于自由度为 $M + 1$ 的 $\chi^2$ 分布的 $p$ 值,并将其除 2(Berkhof and Snijders, 2001),可找到正确的近似 $p$ 值。对于包含多个随机效应的模型,沃贝克和莫莱恩伯(Verbeke and Molenberghs, 2003)针对方差成分推导出一般**单边得分检验**(general one-sided score tests)。在这方面相对不那么技术性的讨论,我们建议参考 Snijders and Bosker(1999)以及 Verbeke and Molenberghs(2000)。

另一种违背正则条件的情况出现在潜类模型中，这种潜类模型通过在 $k$ 个分类的模型上施加简单约束也无法得到 $k-1$ 个分类的模型。例如，将一个类别的概率约束为 0，会使得对应的原点无法辨识。或者，如果将两个类别的原点设为相等，那么则意味着仅有对应的概率总和可以被辨识。因此，在这个情形中，似然比值统计量并不存在 $\chi^2$ 分布（如 Aitkin and Rubin，1985；Titterington et al.，1985；McLachlan and Basford，1988；Everitt，1988）。一个可能的解决方法是使用参数自助法，从 $K-1$ 分类的模型中模拟数据，并同时估计 $K-1$ 分类模型和 $K$ 分类模型以计算似然比值统计量。随后，这些统计量在自助法迭代中的经验分布被用于计算近似显著概率。进一步的讨论和模拟，我们建议参考波宁（Böhning，2000，第 4 章）。

### 8.3.5 置信区间

置信区间可通过将似然比值检验、瓦尔德检验或得分检验进行转化而得到。对于参数 $\beta$，为了构建其 $100(1-\alpha)\%$ 的置信区间，我们必须找到置信下限 $\beta_l$，使得零假设 $\beta=\beta_l$ 及其备择假设 $\beta>\beta_l$ 的单边检验的 $p$ 值等于 $\alpha/2$。类似地，还得找到置信上限。

基于瓦尔德检验的置信区间则为

$$\hat{\beta} \pm z_{1-\alpha/2}\mathrm{SE}(\hat{\beta})$$

其中，$z_{1-\alpha/2}$ 为标准正态分布的 $1-\alpha/2$ 分位数，$\mathrm{SE}(\hat{\beta})$ 为 $\hat{\rho}$ 的估计标准误。

然而，对于基于得分和似然的区间，则需对置信界限进行查找。对于基于似然的区间，需要在给定剖面似然的条件下，对于所关注的参数 $\beta$ 的每一个固定值，估计其他参数的最大似然估计的对数似然。置信界限就是当剖面对数似然为 $\chi^2_{1-\alpha}(1)$ 时 $\beta$ 的数值，并且该数值小于关于全部参数的对数似然最大值，其中 $\chi^2_{1-\alpha}(1)$ 为卡方分布的 $1-\alpha$ 分位数，自由度为 1（在 95% 置信区间下为 3.84）。

在应用章节部分，我们提供参数估计的估计标准误，并在需要的时候，建构基于瓦尔德检验的置信区间。在 9.7 节，使用剖面对数似然法（基于偏差）来计算白靴兔总体规模的置信区间。在 9.4 节的三参数对数概率题器响应模型中，对于猜测参数，我们使用基于剖面对数似然的置信区间。

方差成分的置信区间建构仍然存在一些问题，特别是当估计值接近于 0 时。博塔伊（Bottai，2003）认为置信区域应基于得分检验并使用期望信息。

## 8.4 模型选择：相对拟合标准

竞争模型经常通过相对拟合标准进行对比（如 Jöreskog，1974；Tanaka et al.，1990；Tanaka，1993）。在 5.3 节已经提过，等价模型无法在经验上进行区分，尽管它们可能表示不相同的甚至相互矛盾的实质性过程。对于不等价模型同样如此，尽管对于所分析的数据，它们可能恰好得到相似的拟合。对于这样的情形，模型选择必须按照实质性的或者其他非统计的观点进行。

### 8.4.1 嵌套模型的显著性检验

似然比值检验、瓦尔德检验及得分检验可用于比较嵌套模型。然而，这个方法并不太适用于模型选择，至少有 5 个方面的理由。

其最主要的理由在于，我们很少能够确定哪些模型用来与先验模型做比较。相反，

检验是由数据决定,而这数据还将用于模型评估。换句话说,我们处在“模型生成”的情景之中。很显然,这样的情景并不适合用传统的检验范式,并且总体的模型选择策略的抽样属性也是未知的(如 Freedman,1983)。拒绝传统策略的第二个理由是,条件依赖于某个被选择的模型将忽视模型的不确定性,并将导致标准误的低估(如 Miller,1984)。传统检验方法的第三个问题是,假设检验的效力依赖于样本规模。尽管更多的观测意味着更多的信息,但是将模型选择基于检验标准仍然毫无道理。当考虑一个具有大量观测的情形时,这一点将更为显而易见。在这种情形中,除了那些极其复杂的模型,我们将期望拒绝所有的模型,最后所留下的仅仅是对手头数据做镜像的“模型”。从另一方面来说,如果观测较少,我们期望保留过度简化的模型。第四,必须清楚,所感兴趣的几个模型可能是**非嵌套模型**。因此,在这种情形中,对于拟合度的评估就无法基于传统的检验标准。然而,仍然可使用非嵌套模型的检验,如考克斯(Cox,1961,1962)所提出的方法,以及达维森和麦基依(Davidson and MacKinnon,1993,第 11 章)对相关的检验所做的综述。第五,不少学者认为具有显著性的概率与证据之间经常存在矛盾,即便是对于只存在两个嵌套模型这种不可能的情形(Berger and Sellke,1987;Berger and Delampady,1987)。

然而,值得一提的是,从严格的意义上讲,不应对显著概率过度解释,它应当仅仅被当做是对拟合指标的不太正式的表示(如 Jöreskog,1969,1978)。

### 8.4.2 贝叶斯模型选择,后验比率、贝叶斯因子及贝叶斯信息标准(BIC)

假设我们需要用数据 **D** 来对几个可能的非嵌套的竞争模型进行比较。一种贝叶斯方法为在给定数据的条件下,对模型的后验概率进行比较。

根据贝叶斯定理,给定模型 $M_k$ 的后验概率为

$$\Pr(M_k \mid \mathbf{D}) = \Pr(\mathbf{D} \mid M_k)\Pr(M_k)/\Pr(\mathbf{D})$$

其中,$\Pr(M_k)$ 为模型的后验概率,$\Pr(\mathbf{D}|M_k)$ 为在给定模型的条件下,数据的边缘概率,$\Pr(\mathbf{D})$ 为数据的边缘概率。相对于竞争模型 $M_l$,数据支持模型 $M_k$ 的程度可用 $M_k$ 相对 $M_l$ 的**后验比率**(posterior odds)来评定,即

$$\frac{\Pr(M_k \mid \mathbf{D})}{\Pr(M_l \mid \mathbf{D})} = \left[\frac{\Pr(M_k)}{\Pr(M_l)} \frac{\Pr(\mathbf{D} \mid M_k)}{\Pr(\mathbf{D} \mid M_l)}\right]$$

公式括号中的第一个部分为后验比率,第二个部分为**贝叶斯因子**(Bayes factor)。通常来说,我们没有足够的理由偏好某个模型,而非先验模型,因此,给予各个模型以相等的先验概率以使得后验比率化约为贝叶斯因子。

值得一提的是,很多贝叶斯派研究者不会选择一个单一模型,而是将所关注的参数的推论基于参数的后验分布上,并在模型中求平均值。这种模型平均方法一般将给定数据当做权重,使用模型的近似后验概率来实现(请见 Wasserman,2000)。

在给定模型的条件下,数据的边缘概率通常被称为综合似然,其定义为

$$\Pr(\mathbf{D} \mid M_k) = \int \Pr(\mathbf{D} \mid \boldsymbol{\vartheta}_k, M_k)\Pr(\boldsymbol{\vartheta}_k, M_k)\,\mathrm{d}\boldsymbol{\vartheta}_k \tag{8.8}$$

其中,$\boldsymbol{\vartheta}_k$ 为模型 $M_k$ 的参数。已有学者指出,综合似然严重依赖于先验分布,尽管先验有可能含糊不清(如 Kass and Raftery,1995)。

很少有积分具有解析表达式,因此,已经有各式各样的近似方案被提出来。对于双倍贝叶斯因子来说,最为简单且最常使用的近似为贝叶斯信息标准(BIC)(如 Schwarz,1978),即

$$\text{BIC} = 2[l(\hat{\boldsymbol{\vartheta}}_{M_k} \mid y, \mathbf{X}) - l(\hat{\boldsymbol{\vartheta}}_{M_l} \mid y, \mathbf{X})] - (v_k - v_l)\ln N \quad (8.9)$$

用6.3.3节所介绍的拉普拉斯近似对式(8.8)求积分可推导出该BIC公式。更进一步的近似为将参数的后验众数替换为最大似然估计值，以及将被积函数的对数的海森矩阵替换为对数似然的海森矩阵，也就是假设似然决定先验分布。可以证明，如果使用"单元先验信息"，即其协方差矩阵等于$N^{-1}$乘于费舍尔信息的倒数的多元正态，那么BIC是对贝叶斯因子的很不错的近似。关于BIC的相对不那么技术化的讨论请见Kass and Raftery(1995)及Raftery(1995)。

尽管只是一个较为粗糙的近似，BIC不仅应用于贝叶斯学派中，在频率学派中也广为使用，因为从统计软件的标准输出中很容易获得这个统计量。对于给定模型$M_k$的BIC，在此标记为$\text{BIC}_k$，通常被定义为

$$\text{BIC}_k = -2l(\hat{\boldsymbol{\vartheta}}_{M_k} \mid y, \mathbf{X}) + v_k \ln N \quad (8.10)$$

选择$\text{BIC}_k$值最小的模型。频率学派发现BIC非常具有吸引力，因为相比起似然比值标准，它可以处理非嵌套模型，并且不需要设定模型参数的先验分布。

然而，BIC难以应用于包含潜变量的模型，因为在式(8.10)中，无法确定应该使用哪个"$N$"。例如，对于双层次模型，包括因子、潜类以及结构方程模型，其第一层次的单元为题器，有的使用群组个数$J$，也有的使用第一层次的单元总数$N$。在潜类模型中，流行的方法是使用$J$(如Vermunt and Magidson,2000;Clogg,1995;McCulloch et al.,2002)，而潜类模型程序GLIMMIX(Wedel,2002b)则使用$N$。在结构方程模型中，伯勒恩(Bollen,1989)及拉夫特里(Raftery,1993)使用$N$，但是拉夫特里(Raftery,1995)也建议使用$J$。在多层次回归模型中，BIC很少被用到。霍伊廷克(Hoijtink,2001)在潜类模型中使用贝叶斯因子，以避免BIC近似以及在$J$和$N$之间的取舍。

另一个与潜变量模型相关的是判断自由度(参数的有效个数)。哈奇斯和萨金特(Hodges and Sargent,2001)、伯恩汉姆和安德森(Burnham and Anderson,2002)以及维伊达和布兰查德(Vaida and Blanchard,2004)认为自由度大于模型参数个数(频率学派的观点，排除潜变量)，小于模型参数个数与潜变量个数的和。

**偏差信息标准(DIC)**

斯皮格哈特等(Spiegelhalter et al.,2002)将模型复杂性的测量用"对所估计的残差信息过剩"的概念来表示，定义为

$$d\{\boldsymbol{\vartheta}, \tilde{\boldsymbol{\vartheta}}, \mathbf{y}\} = -2l(\boldsymbol{\vartheta} \mid \mathbf{y}, \mathbf{X}) + 2l(\tilde{\boldsymbol{\vartheta}} \mid \mathbf{y}, \mathbf{X})$$

其中，$\boldsymbol{\vartheta}$为实际参数向量，$\tilde{\boldsymbol{\vartheta}}$为所估计的参数向量。这也可被看作过度拟合程度，因为它表示数据偏离估计参数模型相对于其偏离真实参数模型的程度。斯皮格哈特等(Spiegelhalter et al.,2002)提出将该测量的后验期望，即

$$pD \equiv \text{E}_{\boldsymbol{\vartheta}}(d\{\boldsymbol{\vartheta}, \tilde{\boldsymbol{\vartheta}}, \mathbf{y}\} \mid \mathbf{y}) = \text{E}_{\boldsymbol{\vartheta}}(-2l(\boldsymbol{\vartheta} \mid \mathbf{y}, \mathbf{X}) \mid \mathbf{y}) + 2l(\tilde{\boldsymbol{\vartheta}} \mid \mathbf{y}, \mathbf{X})$$

作为对模型复杂性的贝叶斯测量或"有效参数个数"。他们还提出将$-2$乘以对数似然(上式中的第一项)的后验期望当做拟合的贝叶斯测量，即偏差信息标准(DIC)为

$$\text{DIC} \equiv \text{E}_{\boldsymbol{\vartheta}}(-2l(\boldsymbol{\vartheta} \mid \mathbf{y}, \mathbf{X}) \mid \mathbf{y}) = -2l(\tilde{\boldsymbol{\vartheta}} \mid \mathbf{y}, \mathbf{X}) + p_D \quad (8.11)$$

在多层贝叶斯模型中，斯皮格哈特等(Spiegelhalter et al.,2002)认为如果不设定层级这个建模所关注的焦点(在维伊达和布兰查德(Vaida and Blanchard,2004)的术语中，被称为条件焦点或边缘焦点)，就无法定义模型复杂性。这个焦点确定了完整概率模型(另见6.11节)

$$\Pr(\mathbf{D},\boldsymbol{\zeta},\boldsymbol{\vartheta}_1,\boldsymbol{\vartheta}_2) = \Pr(\mathbf{D} \mid \boldsymbol{\zeta},\boldsymbol{\vartheta}_1)\Pr(\boldsymbol{\vartheta}\boldsymbol{\zeta} \mid \boldsymbol{\vartheta}_2)\Pr(\boldsymbol{\vartheta}_2)\Pr(\boldsymbol{\vartheta}_1)$$

是如何被分解为似然和先验成分。在此，$\mathbf{D}$ 为数据，$\boldsymbol{\zeta}$ 为潜变量，$\boldsymbol{\vartheta}_1$ 为“固定”参数，$\boldsymbol{\vartheta}_2$ 为超参数。条件(群组特定的)焦点相当于将似然定义为条件依赖于潜变量和“固定”参数，$\Pr(\mathbf{D}|\boldsymbol{\zeta},\boldsymbol{\vartheta}_1)$，以及将先验定义为关于超参数的边缘，即

$$\Pr(\boldsymbol{\vartheta}_1)\Pr(\boldsymbol{\zeta}) = \Pr(\boldsymbol{\vartheta}_1)\int\Pr(\boldsymbol{\zeta} \mid \boldsymbol{\vartheta}_2)\Pr(\boldsymbol{\vartheta}_2)\,d\boldsymbol{\vartheta}_2$$

边缘(总体)焦点相当于将似然定义为关于潜变量的边缘

$$\Pr(\mathbf{D} \mid \boldsymbol{\vartheta}_1,\boldsymbol{\vartheta}_2) = \int\Pr(\mathbf{D} \mid \boldsymbol{\zeta},\boldsymbol{\vartheta}_1)\Pr(\boldsymbol{\zeta} \mid \boldsymbol{\vartheta}_2)\,d\boldsymbol{\zeta}$$

以及将先验定义为 $\Pr(\boldsymbol{\vartheta}_1)\Pr(\boldsymbol{\vartheta}_2)$。对于条件焦点来说，有效参数个数显然将大于边缘焦点。

### 8.4.3 赤池信息标准(AIC)

赤池信息标准(AIC)(如 Akaike,1987)或其变体(如 Bozdogan,1987)经常被用于模型选择。

设 $f(\mathbf{y}|\mathbf{X};\boldsymbol{\vartheta})$ 表示对于一个事先设定的模型(即似然)给定参数的条件下，响应的分布，$f^*(\mathbf{y}|\mathbf{X})$ 为真实模型的分布。当把 $f(\mathbf{y}|\mathbf{X};\boldsymbol{\vartheta})$ 当做是 $f^*(\mathbf{y}|\mathbf{X})$ 的近似时，“信息损失”可被定义为

$$I(f,f^*,\boldsymbol{\vartheta}) = \int f^*(y \mid \mathbf{X})[\ln f^*(y \mid \mathbf{X}) - \ln f(y \mid \mathbf{X};\boldsymbol{\vartheta})]\,dy \tag{8.12}$$

这就是两个模型之间的库尔贝克-莱伯勒信息(Kullback-Leibler information，简称 K-L 信息)。如果在设定的模型中，来自真实模型的数据趋向于不可能，那么对于真实模型和近似对数似然之间差异的期望值(关于 $\mathbf{y}|\mathbf{X}$ 的真实分布)将会很大。如果 $f(y|\mathbf{X};\boldsymbol{\vartheta}) = f^*(\mathbf{y}|\mathbf{X})$，该测量值等于 0；否则为正数。

很自然，我们可以将来自其他数据 $\mathbf{y}^\bullet$ 的参数估计 $\hat{\boldsymbol{\vartheta}}(\mathbf{y}^\bullet)$ 代入到该库尔贝克-莱伯勒信息中。式(8.12)的期望值就对 $\mathbf{y}$ 过度采样，而 $\mathbf{y}$ 独立于 $\mathbf{y}^\bullet$ 但来自于同一个分布，因此，有时将此期望值理解为交叉校验测量。对于重复样本 $\mathbf{y}^\bullet|\mathbf{X}$ 的库尔贝克-莱伯勒信息的期望值

$$\int f^*(\mathbf{y}^\bullet)I(f,f^*,\hat{\boldsymbol{\vartheta}}(\mathbf{y}^\bullet \mid \mathbf{X}))\,d\mathbf{y}^\bullet$$

(关于 $\mathbf{y}|\mathbf{X}$ 以及 $\mathbf{y}^\bullet|\mathbf{X}$ 的双重期望值)构成了赤池信息标准的基础。$I(f,f^*,\hat{\boldsymbol{\vartheta}})$ 中的第一项为模型间的常量，因而在模型比较中可以被忽略。赤池(Akaike,1973)证明 2 乘以期望值的第二项可被近似为

$$-2\int f^*(y^\bullet \mid \mathbf{X})\ln f(y \mid \mathbf{X};\hat{\boldsymbol{\vartheta}}(\mathbf{y}^\bullet \mid \mathbf{X}))\,d\mathbf{y}^\bullet \approx -2l(\hat{\boldsymbol{\vartheta}} \mid \mathbf{y},\mathbf{X}) + 2v \equiv \text{AIC}$$

其中，$\hat{\boldsymbol{\vartheta}}$ 为观测数据的参数估计，$v$ 为模型参数个数。$2v$ 用于修正由于将最大对数似然当做其双重期望估计量所带来的偏误。

注意，对于常规线性回归模型，AIC 与马罗斯(Mallows)的 $C_p$ 统计量完全相同。AIC，BIC 以及 DIC 全都可以被看做是对模型复杂性偏差的惩罚。对于 BIC 来说，其惩罚高于 AIC，这样使得更简约的模型更容易被选择。在潜变量模型中，正如在 8.4.2 节对 BIC 所讨论的理由，参数 $v$ 的个数应该如何确定并不清楚，请见维伊达和布兰查德(Vaida and Blanchard,2004)。

关于 AIC 和 BIC 的讨论请见祖奇尼(Zucchini,2000)以及威瑟曼(Wasserman,2000)，

其中后者是从贝叶斯角度进行讨论。近年来,克莱斯肯和约尔特(Claeskens and Hjort, 2003)提出了焦点信息标准(FIC),试图根据所感兴趣的给定参数,选择"最优"模型进行推论。

在11.2节中,对于过度离散的计数数据,我们使用了AIC和BIC来比较非嵌套模型。并且,我们有些主观地选择了将固定参数个数作为$v$,并且将群组个数作为$N$。$N$在该例子中即等于单元个数。

## 8.5 模型充分性:全局绝对拟合标准

在一般模型框架中,模型的一个或多个成分都有可能出现设定错误。对于条件响应模型,所有在广义线性模型中可能出现的设定错误都有可能发生,包括遗漏变量、不恰当的链接函数、不恰当的方差函数及不恰当的分布假设。另外,也有可能出现违背条件独立的假设,将随机回归系数错误地设定为固定系数,以及对因子负荷和测量误差方差施加不恰当的约束的情况。潜变量的结构方程也可能被错误设定,包括遗漏相关观测协变量或潜在协变量,错误地设定线性关系,以及在多层次模型中在不同层次上错误设定干扰项分布。

设定错误显现在所设定的模型对数据拟合不足上面。主要的挑战是区分由于抽样变差性造成的拟合不足和由于使用不恰当的模型所造成的拟合不足,前者不是个问题,而后者恰恰相反。由于可能的设定错误问题的多样性,很显然,对于复杂模型来说模型诊断是一件令人头痛的事情。

一种自然的方法似乎是首先在开始模型辨识之前,就考虑是否存在各种类型的设定错误的证据。两种常见的方法为对设定错误的全局检验以及"拟合指标"考察。尽管某些检验或指标对于一些特殊的假设偏离非常灵敏,但是一般来说,模型和数据之间的不一致有很多原因。

### 8.5.1 检验设定错误

**绝对拟合检验**

绝对拟合检验假设存在一个基准模型,这个基准通常为所给定的变量集合的饱和模型。当响应变量与协变量都为定类变量时,这个饱和模型为无约束多项式模型,在完整的列联表中所有单元格的期望频次等于观测频次(如Bock and Lieberman, 1970; Bock and Aitkin, 1981)。对于多元正态响应变量(不包含协变量),该饱和模型为包含无约束的均值、方差及协方差的多元正态密度。重要的是,绝对拟合检验无法检测遗漏变量,因为该基准模型与所包括的特定变量相关。另外要注意的是,当考虑更一般的模型时,不存在一个绝对标准。

在此所讨论的任意一个相对拟合标准都可用于比较所感兴趣的模型和这个饱和模型。最常使用的是似然比值检验,因为在受约束模型为真的零假设下,它的渐进分布已知。将一个模型与饱和模型之间的对数似然差异乘以2就称为偏差(deviance)。对于定类数据,另外一个可选的统计量为皮尔逊$X^2$。在13.5节中,我们使用偏差来评估潜类模型的绝对拟合程度。在9.4节中的题器响应模型,我们同时使用偏差与皮尔逊$X^2$。对于稀疏列联表来说,这些检验仍然有问题,因为无法依赖于近似结果。在潜特质模型和潜类模型中,Glas(1988)、Reiser(1996)、Reiser and Lin(1999)以及Bartholomew and Leung(2002)提出一系列基于边缘表格的检验,如所有两两组合变量的表格。

如果在对数线性或协方差结构模型中使用绝对拟合(如 Bishop et al.,1975;Fornell,1983),则违背了假设检验的逻辑。在协方差结构模型中,零假设相当于受约束模型,而备择假设相当于经验协方差矩阵。值得一提的是,在这种情形中,研究者希望保留零假设而不是备择假设。因而当零假设不能被拒绝时,保留零假设。很明显,在此情形中,相对于标准的统计检验模式来说,零假设和备择假设的位置互换了一下。对于由于小样本和缺乏统计效力所导致的模型保留,弗奈尔(Fornell,1983)指出了与此相关的问题。此外,或许令人更加感到奇怪的是,观测到的变量之间的弱相关反而增加了保留模型的可能。

**豪斯曼和怀特检验**

豪斯曼(Hausman,1978)设定错误检验考虑两个估计量 $\hat{\boldsymbol{\beta}}$ 与 $\tilde{\boldsymbol{\beta}}$,当模型设定正确时,这两个估计量都为一致估计,而当模型设定错误时,它们收敛在不同的界限上。

例如,对一个随机截距模型的固定回归系数进行估计。如果模型设定正确的话,那么随机截距模型的最大似然估计量 $\hat{\boldsymbol{\beta}}$ 以及固定截距模型的普通最小二乘法估计量 $\hat{\boldsymbol{\beta}}_{FE}$(见 3.6.3 节)都是一致的。然而,如果随机截距与其中一个协变量相关(见 3.2.1 节),则随机截距模型的最大似然估计量不再是一致估计,而普通最小二乘法估计量还是一致估计。因而,这两个估计值的差异说明这个随机截距模型设定错误。

为了将这个想法形式化,豪斯曼提出下述检验统计量:

$$\omega_h = (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{FE})'[\text{Cov}(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{FE})]^{-1}(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{FE})$$

其中,$\text{Cov}(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{FE})$ 为当模型设定正确时,差异的协方差矩阵。该检验统计量服从渐进 $\chi^2$ 分布,其自由度等于 $\text{Cov}(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{FE})$ 的秩。豪斯曼证明:

$$\text{Cov}(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{FE}) = \text{Cov}(\hat{\boldsymbol{\beta}}) - \text{Cov}(\hat{\boldsymbol{\beta}}_{FE})$$

这使得该检验实现起来较为容易,因为只需要这两个估计量的估计协方差矩阵。

尽管这个方法容易实现并且具有潜在用途,但是仍然存在一些限制。首先,与其他拟合检验相同,该检验对于不同类型的设定错误具有灵敏性,这就使得它难以正确确定问题所在。其次,模拟研究表明,对于一般的样本规模来说(如 Long and Trivedi,1993),该检验的效力可能很低。最后,在无限样本中,该检验统计量的抽样分布可能无法满足近似 $\chi^2$ 分布。

众所周知,对于最大似然理论来说,参数估计值的估计协方差矩阵由式(8.4)的夹挤估计量确定。如果模型设定正确,那么夹挤估计量化约为信息矩阵 $-\mathbf{H}^{-1}$ 的逆矩阵。对于豪斯曼检验来说,这两个估计量之间的差异表明模型设定错误。因而,怀特(White,1982)的信息矩阵检验使用下面这个检验统计量来比较这两个协方差矩阵:

$$\mathbf{d}'\hat{\mathbf{C}}^{-1}\mathbf{d}$$

其中,$\mathbf{d}$ 为估计协方差矩阵元素子集之间差异的向量,其所关联的协方差矩阵为 $\hat{\mathbf{C}}$。该检验与豪斯曼检验一样,对不同类型的设定错误具有灵敏性。另外,它要求 $\mathbf{C}$ 的估计量,因而比较难以实现。

## 8.5.2 优度拟合指标

正如前面所述,在绝对拟合检验中,零假设和备择假设的角色被互换了,研究者希望拒绝关于模型为真的零假设"失败"。稀奇的是,过高的效力反而成为一个问题。

对于这个问题的一个处理方法是使用所谓的优度拟合指标(GFI),典型的例子如本

特勒和博奈特(Bentler and Bonett,1980)针对线性结构方程模型所提出的增量拟合指标 $\Delta$,即

$$\Delta = \frac{F_b - F_m}{F_b}$$

其中,$F_m$ 和 $F_b$ 为用于估计参数(见6.2节)的拟合函数的值,它们分别针对所感兴趣的模型和基准模型。这个指标可被理解为基准模型和所感兴趣的模型在拟合函数上削减的比例。在协方差结构模型中,对基准模型的选择一般为响应变量之间相互独立的模型。请注意,在线性回归中的多重相关系数的平方 $R^2$ 也可用同样的方式定义,即拟合函数为残差平方和,并且基准模型不包含协变量。

拟合优度指标的一个吸引力在于它们一般的取值范围为从0到1,当数值在0.90以上时,通常被认为是"好的拟合"。该方法所存在的主要问题为需要有敏锐的模型选择能力。这么多种优度拟合指标并没有真正解决问题,请见 Bollen(1989),Marsh et al.(1988)以及 Mulaik et al.(1989)的调查以及标准软件的输出。

对基准模型或零模型的选择非常重要(Sobel and Bohrnstedt,1985)。有人认为,将拟合优度定义为一个给定模型相对于一个已知不充分的基准模型毫无意义。例如,在测量模型和纵贯性模型中,响应变量相互独立的基准模型本来就是错误的。因而毫不奇怪,在这种情形下模型能获得很高的拟合优度指标。

另外,基准模型对数据拟合不佳的程度,严重依赖于真实模型参数的量值。例如,估计一个简单平行测量模型。如果真实模型为一个同属测量模型(在因子负荷上以及题器间的测量误差方差上存在足够的变差),因为零模型拟合地非常差,或者换句话说是因为题器的信度很高,所以拟合指标数值将会很高。然而,如果真实模型为信度较低的平行测量模型,那么拟合指标数值将会很低,即使我们估计的是正确的模型。简单来说,如果一个简单线性回归模型的关系实际上是带有一定线性趋势的二次方,对其估也能得到高的 $R^2$;而如果模型为真但斜率低时(相对于总体方差),只能得到低的 $R^2$。

戈德伯格(Goldberger,1991:177)将此总结为:

> *在经典回归模型中,没有任何东西要求 $R^2$ 必须很高。因而,高的 $R^2$ 值并不表明支持该模型,而低的 $R^2$ 值也不表明否定该模型。*

因而,或许拟合指标应该被形容为"决定系数",而这也是经常用来形容 $R^2$ 的。

### 8.5.3 近似误差

在协方差结构模型中,科达克和亨莱(Cudeck and Henly,1991)分析了4种不同的协方差矩阵之间的矛盾。当估计一个模型时,拟合函数 $\hat{F} \equiv F(\hat{\boldsymbol{\Sigma}}, \mathbf{S})$ 将 $n \times n$ 的样本协方差矩阵 $\mathbf{S}$ 与所估计的模型所暗含的基于 $v$ 参数的协方差矩阵 $\hat{\boldsymbol{\Sigma}}$ 相比较。如果在总体中估计该模型,类似的矩阵就为真实的协方差矩阵 $\boldsymbol{\Sigma}_0$ 以及近似模型所暗含的 $\tilde{\boldsymbol{\Sigma}}$。由于总体矩阵未知,**因近似而造成的差异也未知**,其定义为 $\mathrm{F}_0 \equiv \mathrm{F}(\tilde{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}_0)$。

可以证明,由于近似 $\mathbf{F}_0$,**样本差异函数** $\hat{\mathbf{F}}$ 为差异的有偏估计量。一个偏差较少的估计量为 $\hat{\mathbf{F}}_0 = \hat{\mathbf{F}} - \mathbf{J}^{-1}\mathbf{d}$(见 McDonald,1989),其中 $d$ 为自由度($d = \frac{1}{2}n(n+1) - v$)。如果 $\hat{\mathbf{F}}_0$ 为负,布朗尼和科达克(Browne and Cudeck,1993)建议将其设为0。

为了惩罚模型复杂性,斯迭戈(Steiger,1990)提出了近似误差均方根(RMSEA),其定

义为

$$\widehat{\text{RMSEA}} = \sqrt{\frac{\hat{F}_0}{d}}$$

布朗尼和科达克(Browne and Cudeck,1993:144)认为:“我们认为 RMSEA 的值在 0.08 左右或更低就表明可接受近似误差,而如果 RMSEA 大于 0.1 则不应使用这个模型。”他们还证明可以构建 RMSEA 的置信区间以及检验。进一步的讨论请参考布朗尼和阿明戈尔(Browne and Arminger,1995)。

### 8.5.4 交叉校验

使用诸如优度拟合这样的指标,并用模型建构所用的同一个数据来校验一个模型将导致过于乐观的评估结果。在模型选择中使用相对拟合标准也存在同样的问题。对于这个问题的一种补偿方案为使用交叉校验法,可用几种不同的方式来实现。

其中一个显而易见的方法为将样本随机地拆分成测定样本和验证样本,测定样本用于估计候选模型,而验证样本用于检验模型(见 10.3.3 节的例子)。然而,这个方法相当浪费,因为在测定和验证过程中,数据的一大部分被忽略掉了。

另一个方法为重复估计模型,即每次从数据中删除一个单元。当单元 $i$ 被忽略时,所得到的估计被用于计算该单元对差异测量的贡献(如 Stone,1974;Geisser,1975)。很显然,只有当差异测量为关于个体单元贡献的函数时,才能使用该方法。例如,它不适用于在结构方程模型中常用的基于协方差矩阵的差异测量(见 6.2 节)。

交叉校验的一般方法为**再抽样**(resampling)技术(如 Efron and Tibshirani,1993,第 17 章),如自助法。简单来说,就是在每个自助法样本中估计参数,并从原始数据中计算拟合优度指标。然后,通过自助法样本所得到的指标均值就作为交叉校验的测量。

在某些情形中,基于来自单一样本的数据也可得到交叉校验指标的期望值估计(通过重复校验以及验证样本)。这样的例子包括在多元回归中的修正 $R^2$ 以及布朗尼和科达克(Browne and Cudeck,1989)针对协方差结构模型所提出的期望交叉校验指标(ECVI);另见 Browne and Cudeck(1993)以及 Browne(2000)。

## 8.6 模型诊断:局部绝对拟合标准

在发现模型存在某些问题的迹象之后,接下来的步骤就为如何诊断问题。模型诊断为用于发现或多或少违背某些模型假设的步骤。第一步通常为推导出一些反应特定的模型偏离的统计量,如残差。下一步为“检查”该统计量,并设计用于检测问题的形式或非形式的标准。该步骤包含很多方法,包括画图、特殊的阈值、基于理论分布或者基于模拟(如后验估计检查)的形式检验。最后,我们将决定是否采取行动以及采取何种行动。

另一种方法为通过详细展开模型来直接分析设定错误的具体形式。从而,诊断的问题就转化为模型选择的问题。

根据所用模型类型以及响应的类型,可定义不同类型的残差。在某些例子中,残差可能被定义为模型所暗含的描述统计量与“观测到的”描述统计量之间的差异(见 8.6.1 节)。更一般地说,残差经常被定义为针对个体单元。然而,在潜变量模型中,在不同层次上都含有“单元”,因而在不同层次上都可定义残差。与线性回归中的残差类似,处于高层次上的残差是对模型中的潜干扰或残差的预报。然而,原则上来说,根据所分析的

模型背离的类型,任意一个差异函数都可用在任意层次上面。当在不同层次上都存在残差时,按照什么顺序来分析并不太清楚。斯奈德和伯克霍夫(Snijders and Berkhof,2004)提出从下至上的角度,即从第一层次开始;而朗福特和路易斯(Langford and Lewis,1998)则提出从上至下的角度。在线性多层次模型中支持从下至上角度的一个论据为,只有在这种情形中,才可能定义一个不受第二层次残差干扰的第一层次残差,反之则不能(Hilden-Minton,1995)。

### 8.6.1 描述统计量残差

在协方差分析中,残差一般被定义为模型预测与经验协方差或相关之间的差异(如Costner and Schoenberg,1973)。在列联表中,显而易见的残差是基于模型预测与观测到的单元格频次之间的差异,通常对其进行标准化(见 Agresti,2002)。除了分析完整的列联表,在所有类型的成对(边缘)表格中,对残差或拟合优度进行分析也非常简单。如,在潜变量模型中,这些成对表格可能表明某对变量违背了条件独立假设(如 Glas,1988;Vermunt and Magidson,2000,附录)。

### 8.6.2 较高层次上的残差

在多层次数据中,处于不同层次上的群组的理论残差就是其对应的干扰项 $\boldsymbol{\zeta}^{(l)}$。在第 7 章所讨论的计分方法可用于预测或估计这些残差,最常用的方法为经验贝叶斯。对于线性模型来说,使用最大似然估计量(在线性混合模型中被称为 OLS)可能更好,因为正如斯奈德和伯克霍夫(Snijders and Berkhof,2004)指出,它们不太依赖于模型假设。然而,如果第一层次上的模型假设不成立,沃特诺等(Waternaux et al.,1989)则建议使用经验贝叶斯。另一个支持经验贝叶斯的观点认为该方法可用于定类和离散响应变量,而最大似然方法对此则存在问题。举个二项响应变量的例子,如果群组的所有响应都等于 1 或都等于 0(正确地说,就是可能的异常值),则无法得到估计。

在诊断中,经验贝叶斯估计的恰当的标准误为非条件抽样标准差,因为在对模型重复抽样的条件下它反映了估计残差的变差。这个"诊断标准误"被朗日和莱恩(Lange and Ryan,1989)、戈德斯坦恩(Goldstein,2003)、朗福特和路易斯(Langford and Lewis,1998)以及路易斯和朗福特(Lewis and Langford,2001)用于线性混合模型中。在 11.3.3 节中,对于计数数据的经验贝叶斯预测,我们使用近似抽样标准差(见式(7.8),7.3.2 节)来定义标准化残差。

正如在 7.3.2 节所说的,所估计残差不仅在相同的层次内相互相关,在层次之间也是如此,尽管我们通常没有意识到这一点。另外,预测依赖于处于同一群组中的潜变量的真值,见 7.3.1 节。由于这些原因,使得难以确定问题的来源。如,在一个包含随机截距的双层次模型中,由于估计截距和真实斜率之间的相关,真实斜率高而真实截距一般的群组将会出现高的估计截距。同样的,处于第二层次的单元将会有高的估计残差,因为处于同一第三层次内的其他第二层次单元的真实残差很高。

与分别对每个顶层单元去计算不同残差的方法相反,更好的方式是使用基于顶层单元的全部残差的差异测量。对于双层次线性混合模型,斯奈德和伯克霍夫(Snijders and Berkhof,2004)定义了一个标准化的第二层次残差,即

$$\widetilde{\boldsymbol{\eta}}_j^{(2)\prime}\left[\mathrm{Cov}_{\mathrm{y}}(\widetilde{\boldsymbol{\eta}}_j^{(2)} \mid \mathbf{X}_j;\hat{\boldsymbol{\vartheta}})\right]^{-1}\widetilde{\boldsymbol{\eta}}_j^{(2)}$$

其中,协方差矩阵为在 7.3.2 节中所讨论的边缘抽样协方差矩阵。他们证明,这个残差

等于线性混合模型用最大似然(OLS)所估计的残差。斯奈德和伯克霍夫(Snijders and Berkhof,2004)将估计的协方差矩阵当做已知,并指出该残差具有一个近似卡方分布,其自由度为 $M$($M$ 为在第二层次上的潜变量的个数)。

对于多层次模型中的顶层群组,当群组被用一种特定的方式包含进来时,例如在群组中包含一个单独的固定截距,在这种情况下,另一种可能的残差为对数似然差异(Longford,2001)。或者,我们也可以使用顶层群组的对数似然贡献,这类似于斯奈德和伯克霍夫(Snijders and Berkhof,2004)在双层次线性混合模型中所使用的多元残差。

### 8.6.3 评估潜变量贡献

在响应变量和潜变量都为正态分布的模型中,残差的经验贝叶斯预测的经验分布通常被用于评估潜变量的正态性。朗日和莱恩(Lange and Ryan,1989)按照这个思路,在线性混合模型中作了一个标准化的潜变量预测线性组合加权正态分位点图。对于包含非正态响应的模型,经验贝叶斯预测不应被用于评估正态性假设,因为在这个情形下,经验贝叶斯预测的抽样分布未知。然而,戈德斯坦恩(Goldstein,2003:100)在二项响应模型中仍然用了这个诊断方法。

即使对于线性混合模型来说,上面这个方法也是有问题的,这是由于在某种程度上说,输入决定了输出。特别是,如果在给定潜变量 $g(\mathbf{y}|\boldsymbol{\zeta},\mathbf{X})$ 的条件下,潜变量的多元正态先验分布 $h(\boldsymbol{\zeta})$“决定”条件响应分布,那么无论真实分布 $\mathrm{h}(\boldsymbol{\zeta})$ 是否能导致经验贝叶斯预测的正态抽样分布,潜变量的后验分布 $\omega(\boldsymbol{\zeta}|\mathbf{y},\mathbf{X})$ 都将呈现为正态。沃贝克和拉沙弗里(Verbeke and Lesaffre,1996)用模拟的方法证明了这个问题。尽管真实潜变量分布是两个单独的正态密度的混合,但是由于收缩,潜变量的后验分布(错误地假设为正态潜变量分布)仍然呈现正态。由于这些缘故,使用经验贝叶斯预测通常难以发现对正态性的偏离。

对此问题的一个解决方法为使用残差的最大似然估计,因为它们仅仅依赖于 $g(\mathbf{y}|\boldsymbol{\zeta},\mathbf{X})$,并且不受所假设的潜变量分布的影响。另一种方法是释放潜变量的正态性假设。沃贝克和拉沙弗里(Verbeke and Lesaffre,1996)提出设定一个成分个数已知的正态密度混合。我们倾向于使用非参最大似然估计(NPMLE),因为这个半参数方法并不要求任何潜变量分布假设。拉贝-赫斯基思等(Rabe-Hesketh et al.,2003a)证明,如果正态性成立但不像当真实潜变量分布偏斜时那样受收缩影响,那么基于 NPMLE 的经验贝叶斯预测与基于正态性假设的预测在实际上无法区分。在 11.3.3 节的纵贯性计数数据中,使用了这个方法。

为了比较 NPMLE 模型(包含 $C$ 个质量)与假设为正态的常规模型,艾特金等(Atikin et al.,2004)提出一种似然比值检验方法。对于常规模型,为了使得模型可被看做是嵌套于半参数模型(其原点和概率被约束为等于求积公式的原点和权重),使用 $C$ 点高斯-埃尔米特积分。这个方法所存在的一个潜在问题是,如果 $C$ 很小,那么积分近似结果可能很差。

### 8.6.4 第一层次残差

对于线性回归模型,标准化的残差被定义为

$$\frac{y_i - \hat{\mu}_i}{\hat{\sigma}}$$

其中,$\hat{\sigma}$ 为所估计的残差标准差。

单元的**删除残差**(deletion residual)为当该单元被忽略时,根据参数估计从样本中所得到的残差。其思路是一个异常值可能会导致对残差标准差的过高估计,从而使得标准化的残差太小。对于线性混合模型的删除残差,朗福特和路易斯(Langford and Lewis,1998)给出了表达式。

对于广义线性模型,最常用的残差为皮尔逊、偏差以及安斯康姆残差,它们在伯努利分布和泊松分布中的公式如表 8.1 所示。

**表 8.1　皮尔逊、偏差以及安斯康姆残差**

| | 伯努利分布 |
|---|---|
| 皮尔逊 | $\dfrac{y_i-\hat{\mu}_i}{\sqrt{\hat{\mu}_i(1-\hat{\mu}_i)}}$ |
| 偏差 | $\text{sign}(y_i-\hat{\mu}_i)\sqrt{-2\ln(1-\hat{\mu}_i)}$　if $y_i=0$ |
| | $\text{sign}(y_i-\hat{\mu}_i)\sqrt{-2\ln(\hat{\mu}_i)}$　if $y_i=1$ |
| 安斯康姆残差 | $\dfrac{B(y_i,2/3,2/3)-B(\hat{\mu}_i,2/3,2/3)}{[\hat{\mu}_i(1-\hat{\mu}_i)]^{1/6}}, B(z,a,b)=\int_0^z t^{a-1}(1-t)^{b-1}\mathrm{d}t$ |
| | 泊松分布 |
| 皮尔逊 | $\dfrac{y_i-\hat{\mu}_i}{\sqrt{\hat{\mu}_i}}$ |
| 偏差 | $\text{sign}(y_i-\hat{\mu}_i)\sqrt{2\hat{\mu}_i}$　if $y_i=0$ |
| | $\text{sign}(y_i-\hat{\mu}_i)\sqrt{2[y_i\ln(y_i/\hat{\mu}_i)-(y_i-\hat{\mu}_i)]}$　if $y_i\neq 0$ |
| 安斯康姆残差 | $\dfrac{1.5(y_i^{2/3}-\hat{\mu}_i^{2/3})}{\hat{\mu}_i^{1/6}}$ |

对于包含潜响应公式的模型,如正态概率模型和对数概率模型,阿尔伯特和奇布(Albert and Chib,1995,1997)以及葛尔曼等(Gelman et al.,2000)在完整贝叶斯框架下,使用"潜在数据残差"$y^*-\mu$。与此类似,在频率学派中,"**广义残差**(generalized residuals)"被定义为在给定观测响应 $y$ 的条件下,潜残差的条件期望(如 Gourieroux et al.,1987a;Chesher and Irish,1987),而"模拟残差(simulated residuals)"则是在给定观测响应的条件下,从潜残差的后验分布中进行抽取而得到(Gourieroux et al.,1987b)。

如果潜变量已知,我们可以简单地将其数值代入线性估计量,对于常规残差公式使用$\hat{\mu}=g^{-1}(\hat{v})$,以获得第一层次的残差。但是,由于潜变量的数值未知,如何来定义残差是不确定的。在线性混合模型中,潜变量的经验贝叶斯预测经常被替换掉(如 Langford and Lewis,1998),从而得到残差的后验均值。对于包含潜变量的线性因子和结构方程模型,为了定义题器的残差,伯勒恩和阿明戈尔(Bollen and Arminger,1991)使用了经验贝叶斯估计量(回归法)或者最大似然估计量(Bartlett 法)。为了确认单元的异常题器,他们采用对应的抽样标准偏差将残差标准化。

对于包含非线性链接函数的模型,将潜变量替换为经验贝叶斯预测并无法得到残差的后验均值。在贝叶斯框架中,戴伊等(Dey et al.,1998)以及阿尔伯特和奇布(Albert and Chib,1995)使用给定观测响应条件下的原始残差(未标准化的)$y-\mu$ 的后验分布,另外,

阿尔伯特和奇布还使用给定观测响应条件下的潜在数据残差 $y^* - \mu$ 的后验分布。令人意外的是，在频率学派中，对于包含非正态响应的潜变量模型，几乎没有关于第一层次残差的研究。

对于线性混合模型，海尔顿-明顿（Hilden-Minton，1995）提出对于每个群组估计一个单独的模型，以定义不受第二层次残差所"混淆"的第一层次残差，另见斯奈德和鲍斯克的文章（Snijders and Bosker，1999）。

## 8.6.5 辨识异常值

我们设某个异常值为看起来与所设定的模型不一致的某个单元或群组，这也即是假设大多数单元或群组与模型相一致。由于我们不知道真实残差，异常值检测就必须基于所估计的残差以及它们的抽样分布。如果残差的抽样分布是正态时，如包含正态潜变量和响应的"线性情形"，那么当残差超过某个正态分位点时，它就可被当做是一个异常值。对于其他类型的响应变量，一般来说抽样分布未知，可以使用模拟来得到一个参照分布。

我们用 $T_j$ 来表示群组（或单元）$j$ 的一个残差或差异统计量。一种自然而然的做法就是将数值最大的统计量 $T_{\max}$ 所属的群组（或单元）标记为可能的异常值。然后将相应的群组 $j^*$ 当做先验，进行检验，即比较 $T_{\max}$ 与 $T_j^*$ 的抽样分布。然而，考虑到之后的选择方法，正确的参照分布应为最大统计量 $T_{\max}$ 的抽样分布。这些步骤很容易通过模拟或参数自助法（如 Longford，2001）进行。在每次迭代 $k$，首先根据模型模拟响应，然后估计参数以及计算统计量 $T_j^k$。最后，用最大统计量 $T_{\max}^k$ 的经验分布来获得 $p$ 值。如果我们使用的参数所基于的数据同时也用于根据模型来模拟响应的话，那么我们就非常理想化地将参数当做已知。为了考虑估计的不确定性因素，朗福特（Longford，2001）提出从参数的估计参数分布中对其进行抽样（来自信息矩阵的含协方差矩阵的多元正态）。

有些讽刺的是，近来关于诊断的显著性检验在贝叶斯学派中越来越普遍（如 Gelman et al.，2003；Marshall and Spiegelhalter，2003）。最常见的方法为后验预测校验（posterior predictive checking），在此方法中，差异统计量 $T$ 的"预测分布（predictive distribution）"被定义为

$$\Pr(T \mid \mathbf{y}^{\text{obs}}) = \int \Pr(T(\mathbf{y}) \mid \mathbf{L}) \Pr(\mathbf{L} \mid \mathbf{y}^{\text{obs}}) \mathrm{d}\mathbf{L}$$

式中，$\Pr(T(\mathbf{y}) \mid \mathbf{L})$ 为给定参数 $\mathbf{L}$ 条件下的 $T(\mathbf{y})$ 的抽样分布，$\Pr(\mathbf{L} \mid \mathbf{y}^{\text{obs}})$ 为参数的后验分布，因而 $\Pr(T)$ 可被宽泛地解释为 $T$ 的抽样分布关于参数 $\mathbf{L}$ 后验的均值。用马尔可夫链蒙特卡洛法（MCMC）计算后验预测校验相当简单，见 6.11.5 节。对于每个从 $\mathbf{L}$ 的后验中所抽取的样本 $\mathbf{L}$，在给定 $\mathbf{L}$（形式化的似然）的条件下，对 $\mathbf{y}$ 的条件分布进行抽样。那么 $T(\mathbf{y})$ 的经验分布即为所必需的参照分布。

在潜变量模型或多层次贝叶斯模型中，参数向量 $\mathbf{L}$ 包含潜变量、参数及超参数，$\mathbf{L} = (\boldsymbol{\theta}, \boldsymbol{\zeta})$。在这个情形中，后验预测校验因为过于宽容或保守而遭到批评（如 Dey et al.，1998；Bayarri and Berger，2000；Marshall and Spiegelhalter，2003）。这是因为潜变量 $\boldsymbol{\zeta}_j$ 是在给定群组 $j$ 的响应 $\mathbf{y}_j^{\text{obs}}$ 的条件下，从其后验分布中抽样而得到。根据这些 $\boldsymbol{\zeta}_j$ 模拟而得到的新的响应就与观测到的响应过于接近。这个问题可通过从潜变量的先验分布中抽样来避免，这样既可以反映对群组的抽样，也可以反映对群组内单元的抽样。群组 $j$ 的差异统计量 $T_j$ 的参照分布就成为

$$\Pr(T_j) = \int \Pr(T_j(\mathbf{y}_j) \mid \boldsymbol{\vartheta}, \boldsymbol{\zeta}_j) \Pr(\boldsymbol{\zeta}_j \mid \boldsymbol{\vartheta}) \Pr(\boldsymbol{\vartheta} \mid \mathbf{y}_j^{\text{obs}}) \mathrm{d}\boldsymbol{\vartheta}$$

马歇尔和斯皮格哈特(Marshall and Spiegelhalter,2003)将这种"完整数据混合迭代"方法看做是近似于理想又计算简便的交叉校验方法。交叉校验的参照分布等于上式,唯一不同的地方在于,它所使用的后验 $p(\boldsymbol{\vartheta}|\mathbf{y}^{\text{obs}}_{(-j)})$ 是基于除了群组 $j$ 以外的全部响应。这个想法也很易于被频率学派所接受。

### 8.6.6 影响力诊断

我们已经将异常值定义为看上去与其他数据不一致的那些单元(或群组)。另一种极端类型是对参数估计有很大影响的那些单元,也就是说如果忽略这些单元,将会导致实质性的变化。

单个的顶层群组对参数估计的影响可用该第 $j$ 个群组的库克(Cook)距离来估算,其定义为

$$C_j = -2\mathbf{g}^{j\prime}\mathbf{H}^{-1}\mathbf{g}^{j}$$

其中,$\mathbf{g}^j$ 为群组 $j$ 的得分向量(对数似然贡献的一阶导数),$\mathbf{H}$ 为总体对数似然的海森矩阵。

对于影响力的另一种测量是当删除一个群组时,在参数估计上的变化。设 $\hat{\boldsymbol{\vartheta}}$ 为使用全部样本的参数估计,$\hat{\boldsymbol{\vartheta}}_{(-j)}$ 为当删除群组 $j$ 后的估计。参数 $\vartheta_s$ 的 $\text{DFBETAS}_{s(-j)}$ 就定义为

$$\text{DFBETAS}_{s(-j)} = \frac{\hat{\vartheta}_s - \hat{\vartheta}_{s(-j)}}{\text{SE}(\hat{\vartheta}_s)}$$

每次删除一个群组并对模型重新估计将带来极大的计算负担。在广义线性模型中,普莱吉本(Pregibon,1981)提出将 $\hat{\boldsymbol{\vartheta}}$ 当做初始值,使用单步牛顿-拉夫逊算法来获得对 $\hat{\boldsymbol{\vartheta}}_{(-j)}$ 的近似

$$\hat{\boldsymbol{\vartheta}}^1_{(-j)} = \hat{\boldsymbol{\vartheta}} - \mathbf{H}^{-1}_{(-j)}\mathbf{g}_{(-j)} \tag{8.13}$$

其中,$\mathbf{H}_{(-j)}$ 为不包含群组 $j$ 时的海森矩阵;$\mathbf{g}_{(-j)}$ 为不包含群组 $j$ 时的梯度向量,由于在最大似然估计中,总体梯度向量为 0,它可由下式得到

$$\mathbf{g}_{(-j)} = \sum_{k\neq j}\mathbf{g}^k = -\mathbf{g}^j \tag{8.14}$$

库克距离和使用式(8.13)的单步近似法所得到的 DFBETAS 存在一个简单关系。为了证明这一点,我们首先使用式(8.13)和式(8.14)将得分向量写为

$$\mathbf{g}^j = \mathbf{H}_{(-j)}(\hat{\boldsymbol{\vartheta}} - \hat{\boldsymbol{\vartheta}}^1_{(-j)})$$

接下来,由于在大样本(在大样本中,单个群组对海森矩阵的贡献很小)中,$\mathbf{H}_{(-j)} \approx \mathbf{H}$,并且 $\mathbf{H} = -[\text{Cov}(\hat{\boldsymbol{\vartheta}})]^{-1}$,因此,库克距离可近似为

$$\begin{aligned} C_j &= -2(\hat{\boldsymbol{\vartheta}} - \hat{\boldsymbol{\vartheta}}^1_{(-j)})'\mathbf{H}'_{(-j)}\mathbf{H}^{-1}\mathbf{H}_{(-j)}(\hat{\boldsymbol{\vartheta}} - \hat{\boldsymbol{\vartheta}}^1_{(-j)}) \\ &\approx 2(\hat{\boldsymbol{\vartheta}} - \hat{\boldsymbol{\vartheta}}^1_{(-j)})'\Big[\text{Cov}(\hat{\boldsymbol{\vartheta}})\Big]^{-1}(\hat{\boldsymbol{\vartheta}} - \hat{\boldsymbol{\vartheta}}^1_{(-j)}) \end{aligned} \tag{8.15}$$

式(8.15)在右侧的表达式为 2 倍的普莱吉本(Pregibon,1981)的单步影响诊断(one-step influence diagnostic)。库克距离的切割点通常为参数个数除以观测(在这里为群组)个数的 4 倍。对于 DFBETAS,则通常使用 2 除以观测个数平方根。

勒斯弗和沃贝克(Lesaffre and Verbeke,1998)将库克距离应用到线性混合模型中,欧文斯等(Ouwens et al.,2001)以及向黎明等人(Xiang et al.,2002)则将其应用于广义线性混合模型。欧文斯等(Ouwens et al.,2001)还发展了新的方法来评估第一层次单元的影

响。在11.3.3节中,为了在一个纵贯性数据中辨识出有影响力的主体,我们进行了影响力诊断。

## 8.7 总结与进一步阅读

我们讨论了一些关于模型设定和诊断的方法,但未提供如何操作的指南。其中,一个理由是因为可能不存在一个最优的方案。另一个理由是在这个领域中相关的研究非常稀缺,特别是对于潜变量模型。最后,从非常不同的文献中汇集各方面的建议实在是一件令人畏惧的工作。

尽管在经验研究中,统计建模即使不是最为普及也是非常普及,但是令人奇怪的是,关于统计建模本身的文献却相当缺乏。两篇非常有用的论文来自 Cox(1990)以及 Lehmann(1990)。

利莫尔(Leamer,1978)从贝叶斯的视角对"设定探索(specification searches)"做了一个很有意思的研究,哈瑞尔(Harrell,2001)针对"经验模型"的模型建构做了一些扩展研究。Langford and Lewis(1998)、Snijders and Bosker(1999)、Verbeke and Molenberghs(2000,第9章)以及 Snijders and Berkhof(2004)讨论了线性混合模型的模型建构和诊断策略。伯勒恩和朗(Bollen and Long,1993)讨论了结构方程模型的各种模型建构策略。关于统计推论的书请参见 Cox and Hinkley(1974)、Lindsey(1996)以及 Pawitan(2001)。

关于潜变量模型诊断的研究还处在婴幼时期,特别是对于包含非连续响应变量的模型。最近一本关于线性混合模型(确切地讲是增长曲线模型)诊断的书来自潘建新和方开泰(Pan and Fang,2002)。对于定序线性回归模型,有许多本关于诊断的书,它们对于更一般的模型也有所帮助,包括 Barnett and Lewis(1984)、Belsley et al.(1980)、Cook and Weisberg(1982)以及 Chatterjee and Hadi(1988)。

必须记住的是,即使经过非常仔细的诊断和检查,"最终模型"仍然有可能是设定错误的。如果要对个别参数(如试验效应)进行推论,那么则建议对模型假设的灵敏度进行分析。贝叶斯统计学家有时使用模型平滑,从而得到参数的可信区间,这些可信区间试图考虑模型的不确定性。一个不那么野心勃勃的方法为灵敏度分析(sensitivity analysis),出于对分析推论"稳健性"的考虑,该分析方法对模型的假设进行了修正。

另外要记住的是,标准误趋向于低估不确定性,因为通常来说模型建构所用的数据就是参数估计所用的数据。因为这个缘故,在极端的探索性分析中,都不提供标准误,尽管在一般的探索性分析中它们通常被接受。我们赞同考克斯(Cox,1990:173)的观点,即在这种情形中,标准误应该被理解为下限:

> 绝大多数研究,在任何情况下,都处于验证性-探索性这两种极端中间的某个位置,而一些关于精度的概念,无论其近似程度,如果能够避免过度解释的话,看上去似乎是这个分析的探索性部分希望得到的。将标准误等信息附加在探索性分析(如探索性多元回归)的主要结果之后,似乎通常表明不确定性的最低程度。

# 第 2 部分　应用研究

# 9 二项响应

## 9.1 介绍

本章是关于模型应用部分的第 1 章，我们首先从第 2 章所讨论的经典模型开始，9.2 节讨论随机截距模型，9.3 节讨论潜类模型，9.4 节讨论题器响应模型和 MIMIC 模型。9.5节讨论随机系数模型。随机系数模型在某些场合下应用于元分析。纵贯性数据的一般应用模型将在关于计数数据的 11.3 节予以介绍。

本章接下来的部分将讨论各种不同的非常规模型。在 9.6 节我们将在纵贯性数据模型中引入状态依赖和未观测到的异质性。在 9.7 节我们将使用考虑到未观测到的异质性因素的捕获-再捕获模型来估计总体规模。最后，在 9.8 节我们将考虑多层次题器响应模型。

本章所讨论的这些模型的应用来自于各个不同学科，如教育学、临床医学、流行病学、生物学、经济学、社会学及社会心理学。

## 9.2 儿童呼吸道感染研究：随机截距模型

萨默尔等（Sommer et al.，1983）介绍了一项印度尼西亚学龄前儿童的同期群研究，在该研究中，连续 6 个学期跟踪检查儿童是否患有呼吸道感染。

泽戈和卡里姆（Zeger and Karim，1991）、迪歌等（Diggle et al.，2002）以及其他一些研究者根据其中 275 个儿童的数据子集[1]估计了一个对数概率随机截距模型（logistic random intercept model）。在第 $i$ 个学期，第 $j$ 个儿童患呼吸道感染的模型可以描述为

$$\text{logit}[\Pr(y_{ij} = 1 \mid \mathbf{x}_{ij}, \zeta_j)] = \mathbf{x}'_{ij}\boldsymbol{\beta} + \zeta_j$$

其中

$$\zeta_j \sim \mathbf{N}(0, \psi)$$

泽戈和卡里姆使用了下述协变量：

- [Age]用月数表示的年龄（中心化为 36 个月）。
- [Xero]虚拟变量，是否有干眼症。干眼症是缺乏维生素 A 在眼科上的表征。
- [Cosine]季节的余弦。
- [Sine]季节的正弦。

---

1 数据可从本书网页 gllamm.org/books 或者帕德里克・赫加蒂（Patrick Hegearty）的主页下载，http://faculty.washington.edu/heagerty/Books/AnalysisLongitudinal/xerop.data。

• [Female]虚拟变量,女性。

• [Height]身高相对国家健康统计中心(NCHS)标准身高的百分比(中心化为90%),在这里反映低营养状态。

• [Stunted]虚拟变量,表示身高是否发育迟缓。其定义为身高是否低于同年龄身高的85%。

使用12点自适应积分法对该模型的最大似然估计结果如表9.1所示的左半部分。

**表9.1 对数随机截距模型与GEE模型估计结果**

| | 随机截距模型 | | | 边缘模型(GEE) | | |
|---|---|---|---|---|---|---|
| | Est | (SE) | OR | Est | (SE) | OR |
| $\beta_1$[Age] | -0.034 | (0.007) | 0.967 | -0.032 | (0.006) | 0.969 |
| $\beta_2$[Xero] | 0.624 | (0.480) | 1.867 | 0.621 | (0.441) | 1.861 |
| $\beta_3$[Female] | -0.436 | (0.257) | 0.646 | -0.418 | (0.236) | 0.658 |
| $\beta_4$[Cosine] | -0.594 | (0.174) | 0.552 | -0.568 | (0.170) | 0.567 |
| $\beta_5$[Sine] | -0.165 | (0.174) | 0.848 | -0.162 | (0.146) | 0.850 |
| $\beta_6$[Height] | -0.048 | (0.027) | 0.953 | -0.048 | (0.031) | 0.954 |
| $\beta_7$[Stunted] | 0.202 | (0.441) | 1.224 | 0.149 | (0.411) | 1.161 |
| $\beta_0$[Const] | -2.673 | (0.224) | — | -2.421 | (0.178) | — |
| $\sqrt{\psi}$ | 0.806 | (0.207) | — | — | — | — |
| 对数似然值 | | -334.65 | | | — | |

由表9.1可知,呼吸道感染与年龄[Age]、季节([Cosine],[Sine])及身高[Height]相关。[Age]的优比估计为0.967,即年龄每增长一个月,患呼吸道感染的比率降低3.3%。这也就是在给定随机截距$\zeta_j$(以及其他协变量)的条件下[Age]的条件效应。使用GEE方法估计的边缘效应结果如表9.1所示的右半部分。在这个模型中,相关系数矩阵的结构被设定为可互换的(见3.6.1节的展示3.3A),另外,标准误是基于8.3.3节所介绍的夹挤估计量。在这个例子中,边缘效应及条件效应的估计值比较接近,这是因为随机效应方差的估计值仅有0.65,从而导致衰减因子的值大约为0.9(见4.8.1节)。

如图9.1所示为在第一个学期中,身高发育未受阻滞、未存在干眼症以及身高为平均水平(1.83)的女孩的年龄和呼吸道感染之间的条件关系和总体平均关系。图中3条短画线曲线分别表示$\zeta_j$ = -0.8,0以及0.8下的条件关系(从低到高)。圆点曲线表示用GEE法估计的总体平均关系,而形状与其相吻合的实线表示通过随机截距模型所得到的总体平均曲线,这是通过对随机截距求积分得到的,即

$$E_{\zeta}\left[\frac{\exp(\mathbf{x}'_{ij}\boldsymbol{\beta}+\zeta_j)}{1+\exp(\mathbf{x}'_{ij}\boldsymbol{\beta}+\zeta_j)}\middle|\mathbf{x}_{ij}\right]$$

在这个例子中,随机截距模型与GEE模型意味着呼吸道感染与[Age]之间存在几乎相同的边缘关系,但在一般情况下,并非都需要该假设。

随机截距的方差估计值为$\hat{\psi}=0.650$。通过使用对数概率模型的潜响应表示法(见2.4节),在控制了协变量的情况下,潜响应之间的条件相关系数为$\hat{\psi}/(\hat{\psi}+\frac{\pi^2}{3})=0.165$。

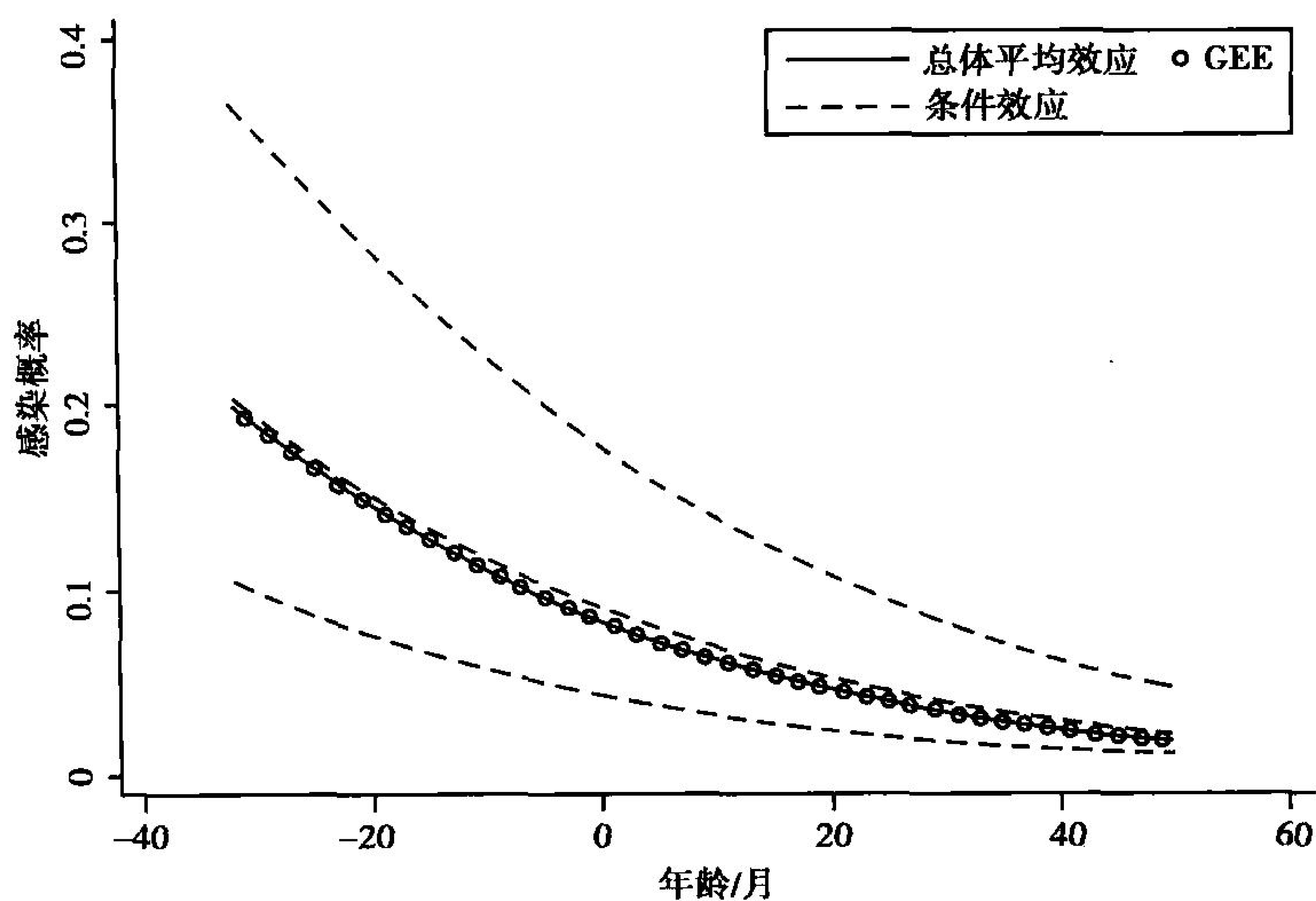



**图 9.1 [Age]的条件效应与总体平均效应**

圆点曲线:GEE 模型估计的总体平均曲线;实线:随机截距模型估计的总体平均曲线;短实线:分别为 -0.8, 0, 0.8 下的条件关系。

观测响应之间的条件相关依赖于协变量的数值。与其相反,GEE 模型假设观测响应之间存在一个皮尔逊相关常量(控制协变量后),其估计值为 0.045。

图 9.2 是儿童的随机截距的经验贝叶斯(见 7.3.1 节)估计的箱图(第 1 个箱图),其分布偏斜并且存在一些极值。然而,通常广义线性混合模型的经验贝叶斯估计的分布并不总是正态的。因此,很难断定这些极值是否值得关注(见 8.6.2 节)。用非形式的方法来评估,我们根据所估计的模型、估计的参数及随机截距,模拟了响应数据(保留数据中的协变量)。重复 3 次,得到图 9.2 中的第 2 到第 4 个箱图。该模拟数据的箱图类似于真实数据,因而似乎没有必要担心那些极值。

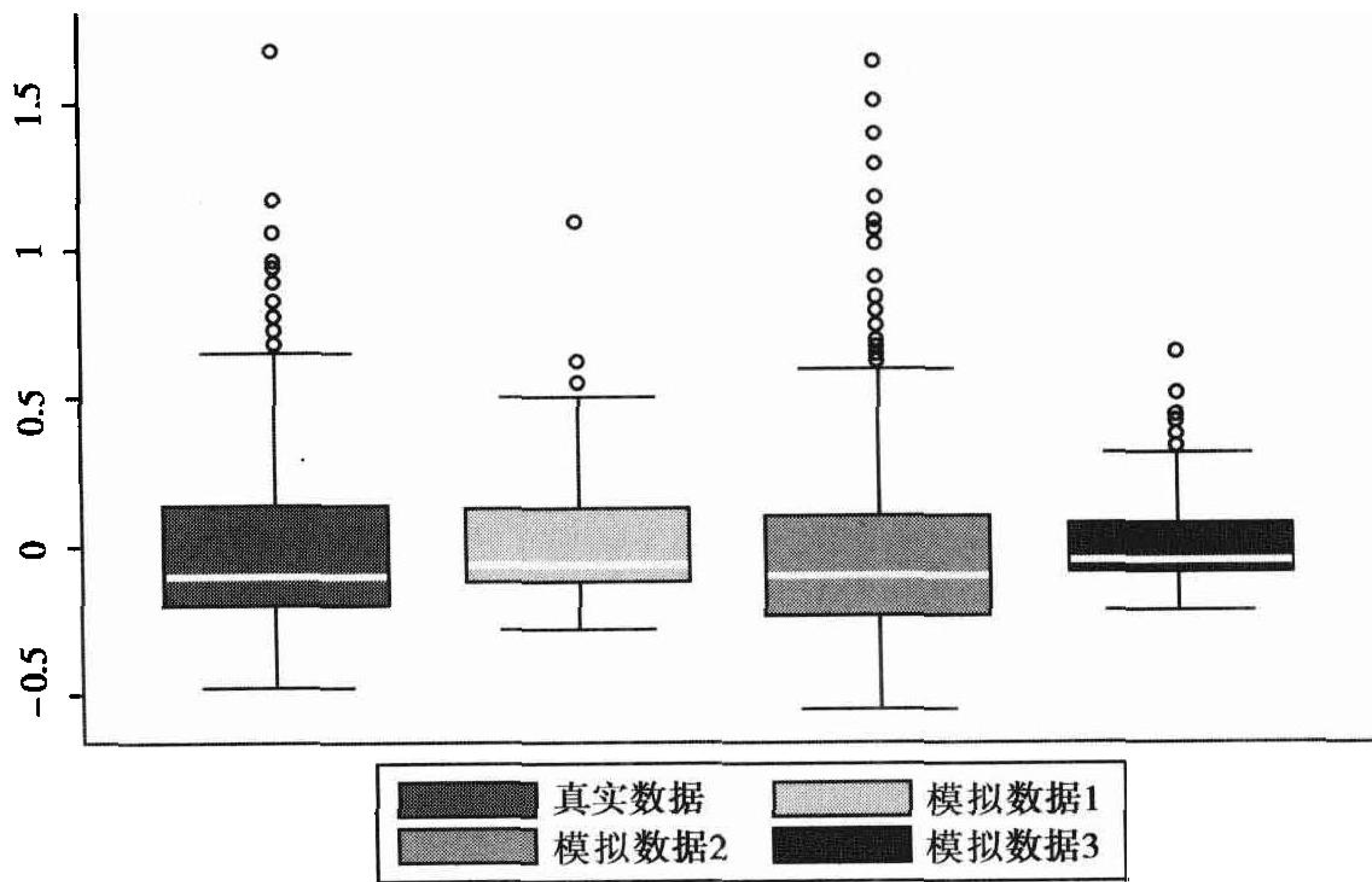



**图 9.2 用经验贝叶斯法对真实数据(第 1 个箱图)及模拟数据(第 2 到第 4 个箱图)估计的随机截距箱图**

如图 9.3 所示为所估计的随机截距及其对应的秩(间隔 5 个秩),误差条形图代表 ±1个后验标准差。

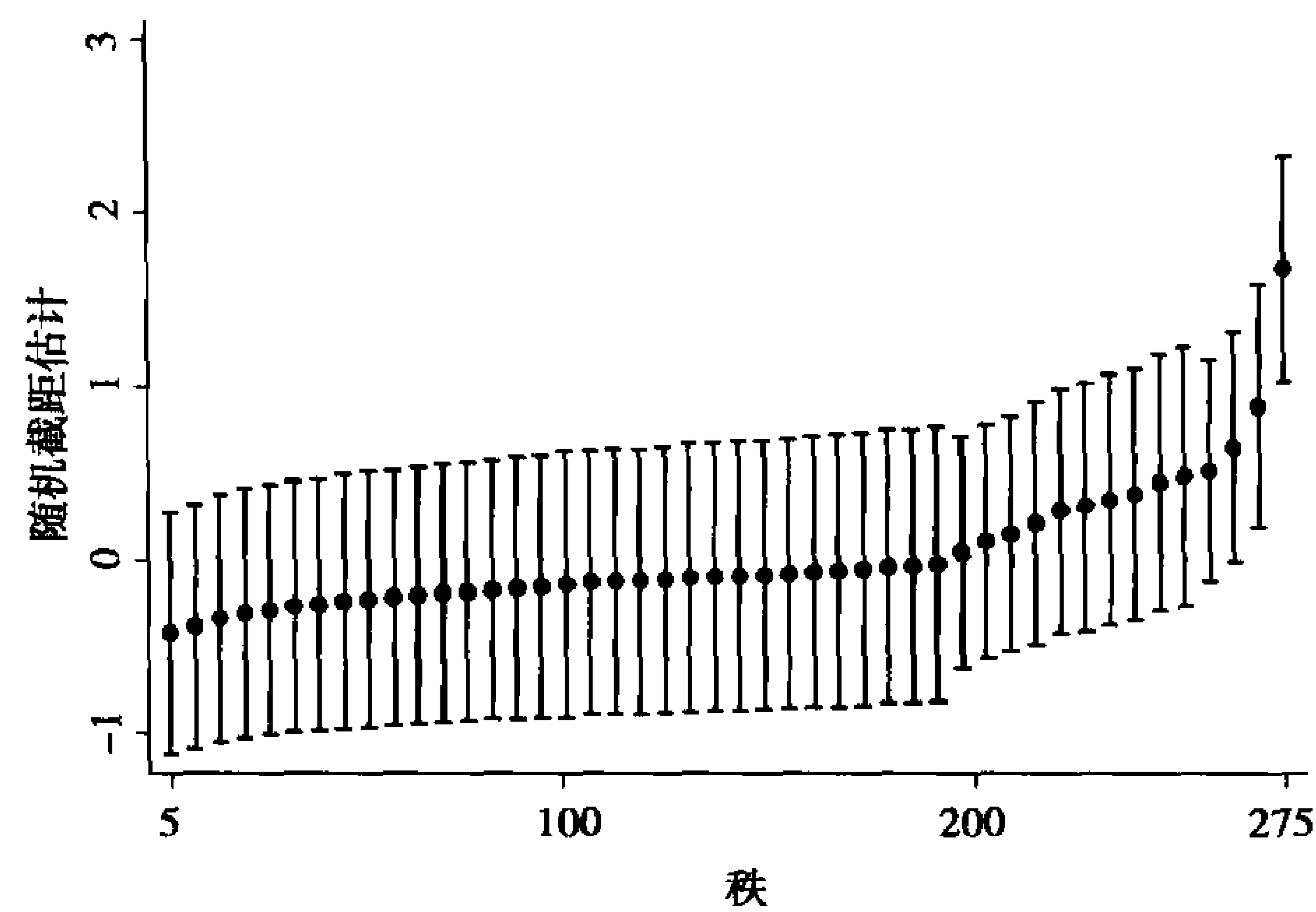



图 9.3　经验贝叶斯估计的随机截距及其秩，误差条形图代表 ±1 个后验标准差

## 9.3　心肌梗塞诊断：潜类模型

林德斯科普夫和林德斯科普夫（Rindskopf and Rindskopf, 1986）分析了纽约市一个冠心病治疗机构的数据[1]，该机构的患者将接受关于心肌梗塞方面的诊断。

总有 94 个患者的数据，每个患者都进行了以下 4 个方面的诊断：

- [Q-wave] 虚拟变量，在 ECG 中是否存在 Q-wave。
- [History] 虚拟变量，是否接受过传统诊疗法。
- [LDH] 虚拟变量，是否存在反转乳酸脱氢酶（flipped LDH）。
- [CPK] 虚拟变量，是否存在肌酸磷酸激酶同工酶（CPK-MB）。

其数据如表 9.2 所示。既然患者只有发生心肌梗塞与没有发生心肌梗塞这两种可能性，故有理由将其分为两个潜在类别。用 $\pi_1$ 表示属于第一个潜在类别的概率，即

$$\text{logit}(\pi_1) = \varrho_0$$

表 9.2　心肌梗塞诊断数据

| [Q-wave] ($i=1$) | [History] ($i=2$) | [LDH] ($i=3$) | [CPK] ($i=4$) | 观测频次 | 期望频次 | 心肌梗塞概率 ($c=2$) |
|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 24 | 21.62 | **1.000** |
| 0 | 1 | 1 | 1 | 5 | 6.63 | **0.992** |
| 1 | 0 | 1 | 1 | 4 | 5.70 | **1.000** |
| 0 | 0 | 1 | 1 | 3 | 1.95 | **0.889** |
| 1 | 1 | 0 | 1 | 3 | 4.50 | **1.000** |
| 0 | 1 | 0 | 1 | 5 | 3.26 | 0.420 |
| 1 | 0 | 0 | 1 | 2 | 1.19 | **1.000** |
| 0 | 0 | 0 | 1 | 7 | 8.16 | 0.044 |

1　数据可从 gllamm.org/books 下载。

续表

| [Q-wave] ($i=1$) | [History] ($i=2$) | [LDH] ($i=3$) | [CPK] ($i=4$) | 观测频次 | 期望频次 | 心肌梗塞概率 ($c=2$) |
|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 0 | 0 | 0.00 | 0.017 |
| 0 | 1 | 1 | 0 | 0 | 0.22 | 0.000 |
| 1 | 0 | 1 | 0 | 0 | 0.00 | 0.001 |
| 0 | 0 | 1 | 0 | 1 | 0.89 | 0.000 |
| 1 | 1 | 0 | 0 | 0 | 0.00 | 0.000 |
| 0 | 1 | 0 | 0 | 7 | 7.78 | 0.000 |
| 1 | 0 | 0 | 0 | 0 | 0.00 | 0.000 |
| 0 | 0 | 0 | 0 | 33 | 32.11 | 0.000 |

资料来源：Rindskopf and Rindskopf(1986)。

如果第二个潜在类别对应于心肌梗塞，那么心肌梗塞的患病率为 $\pi_2=1-\pi_1$，条件响应概率可以设定为

$$\text{logit}[\Pr(y_{ij}=1 \mid c)] = e_{ic}$$

概率 $\Pr(y_{ij}=1 \mid c=2)$ 表示该诊断的灵敏性（正确判断出患病的被测者的可能性），而 $1-\Pr(y_{ij}=1 \mid c=1)$ 表示诊断的特异性（正确判断出没有患病的被测者的可能性）。由于当 $\lambda_1=1$ 时，可以将 $e_{ic}$ 替换为 $\lambda_i e_c$，因此，该模型等价于双分类单因子模型。

参数估计值如表 9.3 所示，对应于[Q-wave]的估计值 $\hat{e}_{11}$ 与对应于[CPK]的估计值 $\hat{e}_{42}$ 分别在正负方向上都很大，其条件响应概率非常接近于 0 或 1，这被称为"**边界解**（boundary solution）"。对应的标准误差也相当大，这是因为当这些极值参数变化时，似然性变化得特别微小，而对于大多数数值来说，所预测的概率仍然保持在接近 0 和 1。例如，当对数概率值为 5 和 20 时，其对应的概率分别为 0.993 与 1.000。

**表 9.3 心肌梗塞估计值**

| 参数 | 第一类别（'NoMI'） | | | 第二类别（'MI'） | | |
|---|---|---|---|---|---|---|
| | Est | (SE) | 概率 | Est | (SE) | 概率 |
| | | | 1 类-特异性 | | | 灵敏性 |
| $e_{1c}$[Q-wave] | -17.58 | (953.49) | 0.00 | 1.19 | (0.42) | 0.77 |
| $e_{2c}$[History] | -1.42 | (0.39) | 0.30 | 1.33 | (0.39) | 0.79 |
| $e_{3c}$[LDH] | -3.59 | (1.01) | 0.03 | 1.57 | (0.47) | 0.83 |
| $e_{4c}$[CPK] | -1.41 | (0.41) | 0.20 | 16.86 | (706.04) | 1.00 |
| | | | 1 类-患病率 | | | 患病率 |
| $\varrho_0$[Cons] | 0.17 | 0.22 | 0.54 | — | — | 0.46 |

根据林德斯科普夫和林德斯科普夫（Rindskopf and Rindskopf，1986），我们不再进一步解释这些参数估计。对于表 9.2 中的每一组测试，期望频次是通过将每个患者的似然分布乘以 94（患者个数）而得到的，即

$$g(\mathbf{y}_j) = \pi_1 \prod_i \Pr(y_{ij} \mid c = 1) + \pi_2 \prod_i \Pr(y_{ij} \mid c = 2)$$

通过在表 9.2 中比较期望频次与观测频次,模型看上去拟合地较好。

根据表 9.3,心肌梗塞的患病率估计为 0.46。[Q-wave]的**特异性**估计为 1,这表明对于没有患心肌梗塞的人将会出现负的结果。[History]的**特异性**估计最低,仅为 0.70。模型灵敏度估计的范围从 0.77([Q-wave])到 1.00([CPK]),即 77% 的心肌梗塞测试结果在[Q-wave]上是正的,在[CPK]上 100% 是正的。

根据用贝叶斯法(见 7.2 节)求得的这 4 组检验结果,我们可进一步得到后验概率(类似于"正向估计值(positive predictive values)"),即

$$\omega(e_2 \mid \mathbf{y}_j;\hat{\boldsymbol{\theta}}) \equiv \Pr(c = 2 \mid \mathbf{y}_j) = \frac{\pi_2 \prod_i \Pr(y_{ij} \mid c = 2)}{\pi_1 \prod_i \Pr(y_{ij} \mid c = 1) + \pi_2 \prod_i \Pr(y_{ij} \mid c = 2)}$$

其概率值如表 9.2 所示的最后 1 列,其中,加粗的字体表示该检验结果的病人属于第二类别的后验概率高于属于第一类别的,因此,根据经验贝叶斯众数分类规则(见 7.4 节),该病人将被诊断为心肌梗塞。对绝大多数病人来说,当它的后验概率接近于 0 或 1 时,该诊断(分类)是相当清楚的。对于每个病人,我们可算出其被错误分类的概率(根据经验贝叶斯众数分类规则)(见 7.13 节)为

$$f_j = 1 - \max_c \omega(e_c \mid \mathbf{y}_j;\hat{\boldsymbol{\theta}})$$

例如,对于 $\mathbf{y}_j = (0,1,0,1)$,该患者将被分类为"未患心肌梗塞",分类错误的概率为 0.42。在这个研究中,具有这么高的分类错误概率的情况仅有这个案例,在全部患者中分类错误的比例预计为 3.26/94 = 3.5%。

根据样本均值 $f_j$,在给定 0.030 的条件下,我们可估计总体的分类错误的比例。如果我们没有检验结果,则只能根据先验概率来诊断患者。那么,每个人都会被诊断为"未患心肌梗塞",因为它的概率($\pi_1 = 0.54$)高于"患心肌梗塞"的概率($\pi_2 = 0.46$)。因此,所估计的分类错误概率就为 0.46。由于已知检验结果,则削减了的分类错误比例为 (0.457 9 − 0.029 6)/0.457 9 = 0.94。如果使用 $f_j$ 的期望值而非样本均值(即使用基于模型的频次期望值而非观测的频次),削减的分类错误比例则为 0.95。引入协变量信息(如年龄与性别)将会进一步提高诊断的精确度,请见 13.5 节协变量潜类模型的例子。

## 9.4 数学推理:题器响应模型

我们将分析美国青年能力数据[1](美国国防部,1982 年),这是一个全国概率样本数据,对美国 16 ~ 23 岁的青年进行能力测试。数据由兵种倾向选择测验中的 4 道数学推理题(编号 8A)构成,样本包括白人与黑人,如表 9.4 所示。米斯勒维(Mislevy, 1985)曾经分析过该数据。

我们首先对题器 $i$、对象 $j$ 估计一个单参数对数概率题器响应模型(one-parameter logistic item response model)(见 3.3.4 节),则

$$\text{logit}[\Pr(y_{ij} = 1 \mid \eta_j)] = \beta_i + \eta_j$$

1 该数据可从 gllamm.org/books 下载。

表 9.4 数学推理数据

| 题器 | | | | 白人男性 | 白人女性 | 黑人男性 | 黑人女性 |
|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | | | | |
| 0 | 0 | 0 | 0 | 23 | 20 | 27 | 29 |
| 0 | 0 | 0 | 1 | 5 | 8 | 5 | 8 |
| 0 | 0 | 1 | 0 | 12 | 14 | 15 | 7 |
| 0 | 0 | 1 | 1 | 2 | 2 | 3 | 3 |
| 0 | 1 | 0 | 0 | 16 | 20 | 16 | 14 |
| 0 | 1 | 0 | 1 | 3 | 5 | 5 | 5 |
| 0 | 1 | 1 | 0 | 6 | 11 | 4 | 6 |
| 0 | 1 | 1 | 1 | 1 | 7 | 3 | 0 |
| 1 | 0 | 0 | 0 | 22 | 23 | 15 | 14 |
| 1 | 0 | 0 | 1 | 6 | 8 | 10 | 10 |
| 1 | 0 | 1 | 0 | 7 | 9 | 8 | 11 |
| 1 | 0 | 1 | 1 | 19 | 6 | 1 | 2 |
| 1 | 1 | 0 | 0 | 21 | 18 | 7 | 19 |
| 1 | 1 | 0 | 1 | 11 | 15 | 9 | 5 |
| 1 | 1 | 1 | 0 | 23 | 20 | 10 | 8 |
| 1 | 1 | 1 | 1 | 86 | 42 | 2 | 4 |
| 合计 | | | | 263 | 228 | 140 | 145 |

资料来源：*Mislevy*(1985)。

其参数估计值如表 9.5 所示，我们可知，所估计的题器难度 $\hat{\beta}_i$ 从题器 1 到题器 4 逐渐增加。该模型假设能力的增加在 4 个题器上的效应(在对数比率的尺度上)都是相同的，可通过使用双参数对数概率题器响应模型去掉这个假设：

$$\text{logit}[\Pr(y_{ij}=1 \mid \eta_j)] = \beta_i + \lambda_i \eta_j$$

其中，为了模型辨识，我们设定 $\lambda_1 = 1$。用 GRC 方程式，该公式可重写为

$$\text{logit}[\Pr(y_{ij}=1 \mid \eta_j)] = \mathbf{d}_i'\boldsymbol{\beta} + \eta_j \mathbf{d}_i'\boldsymbol{\lambda} \tag{9.1}$$

其中，$\mathbf{d}_i$ 是四维向量，其第 $i$ 个元素等于 1，其他元素等于 0。从表 9.5 所显示的参数估计，我们可知，对于题器 2 和题器 3 所估计的判别参数或者说因子负荷 $\hat{\lambda}_i$ 小于其他两个题器。然而，双参数模型并没有比单参数模型拟合得更好。

单参数及双参数题器响应模型所估计的题器特质曲线的定义为

$$\Pr(y_{ij}=1 \mid \eta_j) = \frac{\exp(\nu_{ij})}{1+\exp(\nu_{ij})}$$

其中，线性估计量 $\nu_{ij}$ 在单参数模型中为 $\beta_i + \eta_j$，在双参数模型中为 $\beta_i + \lambda_i \eta_j$。当能力趋向于 $-\infty$ 时，这些曲线逼近于 0。

由于在选择题中，答案是可以被猜测到的，因此，更接近实际情况的模型是三参数对数概率题器响应模型，该模型可表述为

表 9.5　用 20 点自适应积分法估计的单参数、双参数及三参数题器响应模型

| 参数 | 单参数 | | 双参数 | | 三参数 | |
|---|---|---|---|---|---|---|
| | Est | (SE) | Est | (SE) | Est | (SE) |
| 截距 | | | | | | |
| $\beta_1$[题器 1] | 0.58 | (0.10) | 0.64 | (0.12) | -0.05 | (0.18) |
| $\beta_2$[题器 2] | 0.24 | (0.10) | 0.22 | (0.09) | -0.54 | (0.16) |
| $\beta_3$[题器 3] | -0.22 | (0.09) | -0.22 | (0.09) | -1.64 | (0.40) |
| $\beta_4$[题器 4] | -0.59 | (0.10) | -0.63 | (0.11) | -5.07 | (3.73) |
| 因子负荷 | | | | | | |
| $\lambda_1$[题器 1] | 1 | — | 1 | — | 1 | — |
| $\lambda_2$[题器 2] | 1 | — | 0.67 | (0.16) | 0.66 | (0.22) |
| $\lambda_3$[题器 3] | 1 | — | 0.73 | (0.18) | 0.97 | (0.35) |
| $\lambda_4$[题器 4] | 1 | — | 0.93 | (0.23) | 2.42 | (1.92) |
| 猜测参数 | | | | | | |
| $c$ | — | — | — | — | 0.22 | — |
| 方差 $\psi$ | 1.63 | (0.21) | 2.47 | (0.84) | 6.66 | (3.09) |
| 对数似然值 | -2 004.94 | | -2 002.76 | | -1 992.78 | |

$$\Pr(y_{ij}=1\mid\eta_j)=c_i+(1-c_i)\frac{\exp(\nu_{ij})}{1+\exp(\nu_{ij})}$$

其中,$c_i$ 通常被称为“猜测参数”,即一个能力为负无穷的被测者猜测到正确答案的概率。该模型并不符合第 4 章的一般模型框架,这是因为响应模型并不是广义线性模型(在潜变量条件下)。然而,如果我们将猜测参数设置为某个常数,如 0.1,那么该反应模型可以被看做是带有某个复合链接函数的广义线性模型(见 2.3.5 节的式 2.14)。

$$\Pr(y_{ij}=1\mid\eta_j)=0.1g_1^{-1}(1)+0.9g_2^{-1}(\nu_{ij})$$

其中,$g_1$ 是同一性链接函数,$g_2$ 是对数概率链接函数。假设对于所有备选项来说猜测参数相等,我们尝试用 $c$ 的不同数值(从 0 到 0.4,按 0.02 递增),在给定剖面对数似然值(profile log-likelihood)的条件下,控制其他参数使似然值最大化。剖面对数似然值及对应的 $c$ 值如图 9.4 所示,当 $c=0.22$ 时,达到最大值。图 9.4 中用水平短点线表示最大值,近似的 95% 置信极限为当剖面对数似然值小于最大值 3.84/2 时 $c$ 的取值,即 $c$ 的 95% 置信区间为 0.14 ~ 0.28。

当 $c=0.22$ 时,该三参数对数概率题器响应模型的参数估计值如表 9.5 所示。请注意,由于 $c$ 被当做未知,标准误差被低估。该模型的拟合程度明显好于双参数模型。然而,参数估计值并不十分可靠,因为似然曲线看上去显得比较平缓。特别是,估计值 $\hat{\beta}_4$ 与 $\hat{\lambda}_4$ 之间的相关系数估计值为 -0.95。另外,使用不同的初始值将导致非常不同的估计值,但对数似然值却非常近似。因此,经验辨识(见 5.2.5 节)显得较为脆弱。

全部 3 个模型的题器特征曲线如图 9.5 所示。与单参数模型不同,在双参数模型中,与题器 1 和题器 4 相交的曲线的斜率高于题器 2 和题器 3,这明显违背了双重单调性(double monotonicity)。很明显,当能力趋向于 $-\infty$ 时,三参数模型的曲线趋近数值为

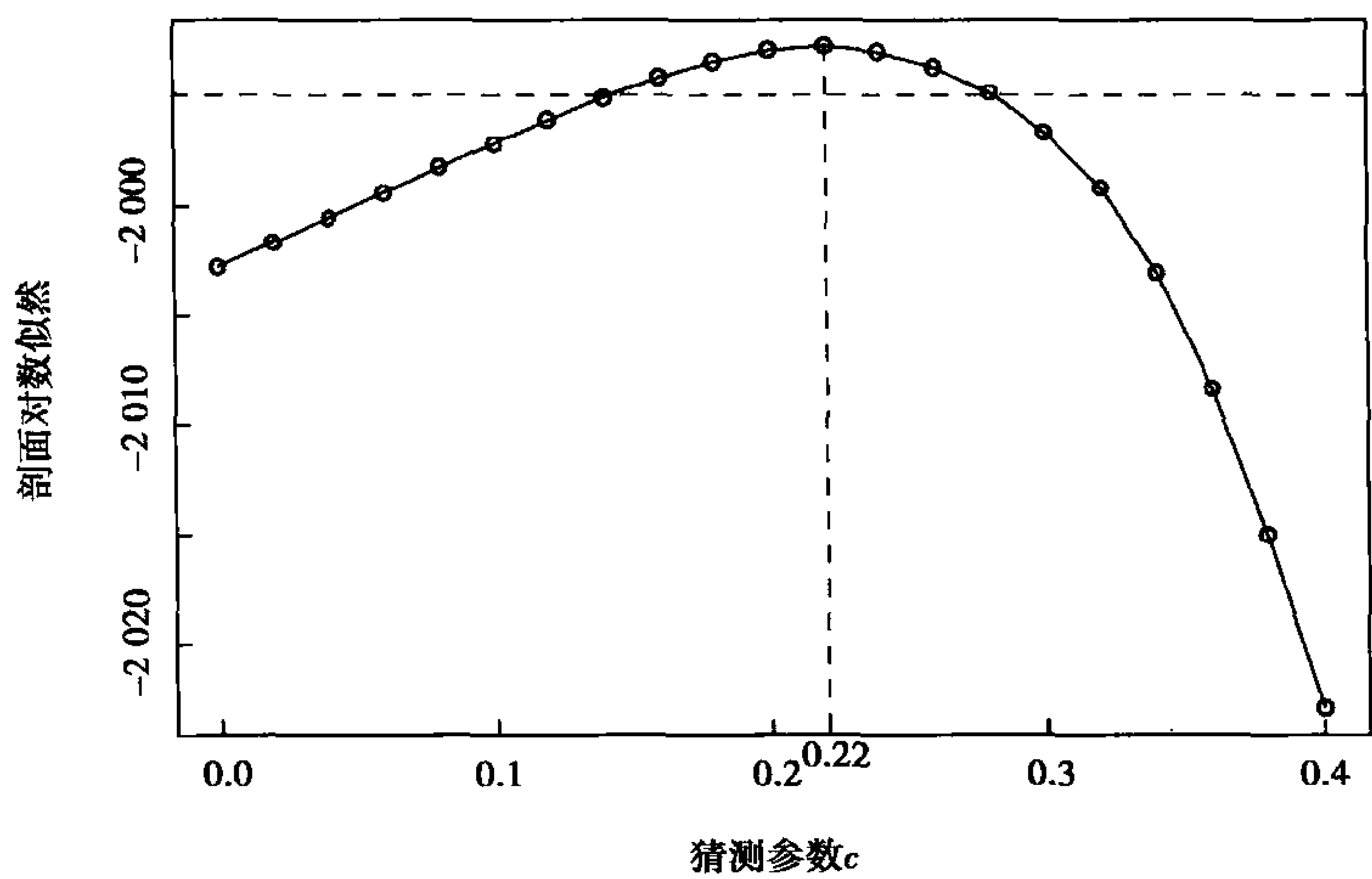



图 9.4 三参数对数概率题器响应模型中猜测参数的剖面对数似然

0. 22的渐近线。

重新回到双参数题器响应模型,我们现在考虑如下协变量:

- [Female] 被测者性别为女性,虚拟变量。
- [Black] 被测者为黑人,虚拟变量。

对于能力 $\eta_j$ 我们可设定一个结构模型,允许能力均值具有组间差异性,即

$$\eta_j = \gamma_0 + \gamma_1 F_j + \gamma_2 B_j + \gamma_3 F_j B_j + \zeta_j$$

其中,$F_j$ 表示[Female],$B_j$ 表示[Black]。既然我们包含了一个常量在该结构模型中,为了模型辨识,我们必须将响应模型中的一个常量固定,并且设置 $\beta_1 = 0$。这种类型的模型也即为 3.5 节所讨论过的 MIMIC 模型,在这种模型中协变量仅通过一个潜变量对响应产生效应。

表 9.6 列出了该模型($M_2$)及不含协变量的模型($M_1$,$\gamma_1 = \gamma_2 = \gamma_3 = 0$)的参数估计。$M_1$ 与表 9.5 中的简单双参数模型是等价的(见 5.3 节)。偏差及皮尔逊卡方统计值如表 9.5 所示,从中我们可知,$M_2$ 拟合得比 $M_1$ 好。由于能力的部分变差性已被协变量所"解释"掉,误差项的方差从 $M_1$ 的2.47 降到 $M_2$ 的 1.88。另外,还存在[Female]与[black]交互效应的证据。在白人中,女性与低能力有较强的关联,而在黑人中,则没有这个现象,男性与女性能力近似。黑人能力均值既低于白人男性,也低于白人女性。

除了协变量通过潜变量对响应变量的非直接效应以外,我们还可研究一下是否存在协变量对响应变量的直接效应。这种效应可被理解为"题器误差"或"题器功能差异(differential item functioning,DIF)",即针对某个题器回答正确的概率是否存在差异,如在具有相同能力的男性和女性之间。如果存在题器误差,则会产生问题,因为它意味着测试将无法公平地评判被测者。巴托罗买(Bartholomew, 1987, 1991)发现,在第一个题器上,黑人女性得分较低。为了研究是否存在这个现象,我们将 $\beta_5 F_j B_j d_{i1}$ 加入式(9.1),以允许能力均值存在组间差异。其参数估计结果如表 9.6 所示的 M3,可知,并没有充分的证据表明黑人女性在题器 1 上的表现有差异。请参见 10.3.3 节关于政治效能的二维定序题器响应模型的题器误差分析。

请注意,根据绝对拟合标准,看起来这些模型没有一个拟合地较好。例如,对 M2 来说,偏差为 111.68,53 个自由度,尽管表格的分布非常稀疏无法依赖于卡方分布来进行检验。

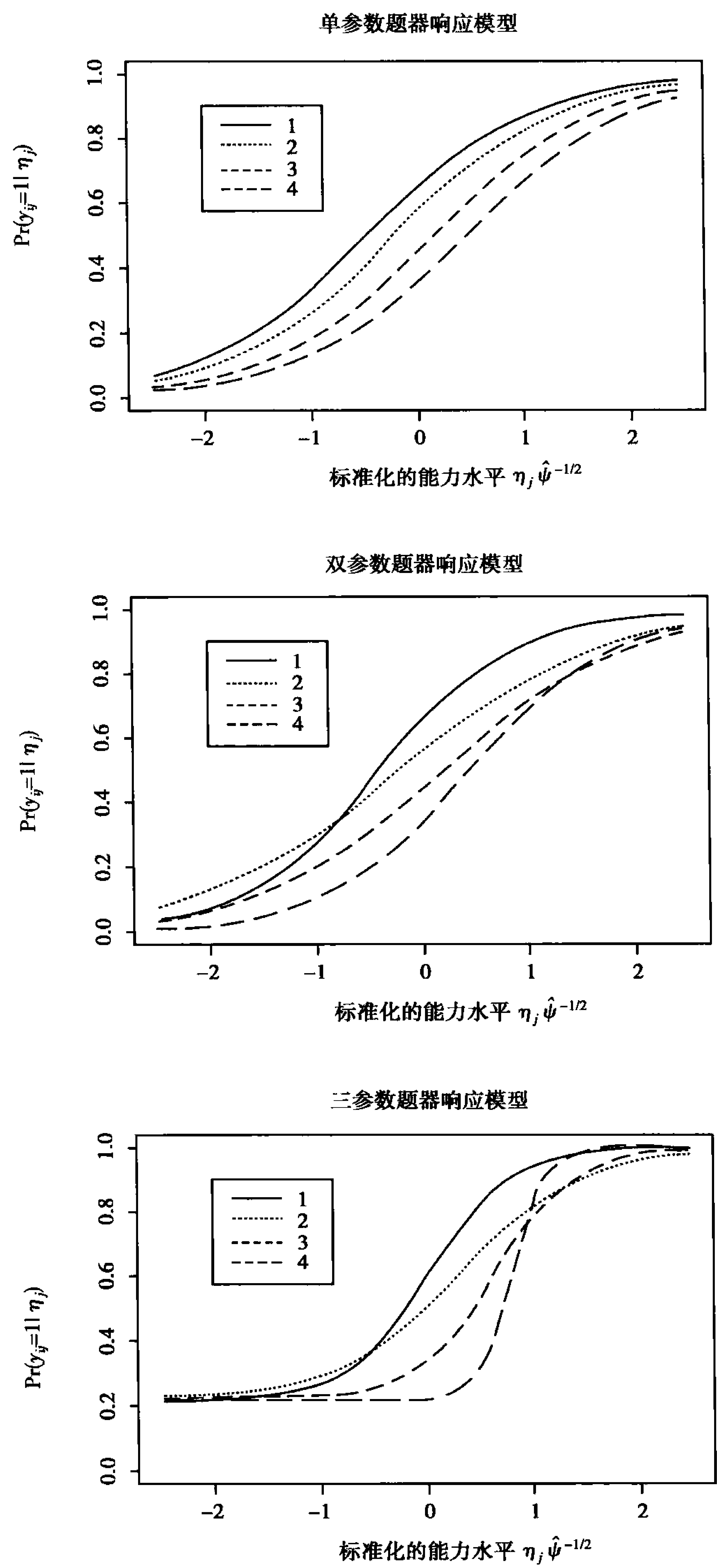



**图 9.5　单参数模型(上)、双参数模型(中)及三参数(下)对数概率题器响应模型的题器特征曲线**

表 9.6 MIMIC 模型估计值

| 参 数 | M1 | | M2 | | M3 | |
|---|---|---|---|---|---|---|
| | Est | (SE) | Est | (SE) | Est | (SE) |
| 截距 | | | | | | |
| $\beta_1$[题器 1] | 0 | — | 0 | — | 0 | — |
| $\beta_2$[题器 2] | -0.21 | (0.12) | -0.22 | (0.12) | -0.13 | (0.13) |
| $\beta_3$[题器 3] | -0.68 | (0.14) | -0.73 | (0.14) | -0.57 | (0.15) |
| $\beta_4$[题器 4] | -1.22 | (0.19) | -1.16 | (0.16) | -1.10 | (0.18) |
| $\beta_5$[题器 1] × | | | | | | |
| [Black] × [Female] | 0 | — | 0 | — | -1.07 | (0.69) |
| 因子负荷 | | | | | | |
| $\lambda_1$[题器 1] | 1 | — | 1 | — | 1 | — |
| $\lambda_2$[题器 2] | 0.67 | (0.16) | 0.69 | (0.15) | 0.64 | (0.17) |
| $\lambda_3$[题器 3] | 0.73 | (0.18) | 0.80 | (0.18) | 0.65 | (0.14) |
| $\lambda_4$[题器 4] | 0.93 | (0.23) | 0.88 | (0.18) | 0.81 | (0.17) |
| 结构模型 | | | | | | |
| $\gamma_0$[Cons] | 0.64 | (0.12) | 1.41 | (0.21) | 1.46 | (0.23) |
| $\gamma_1$[Female] | 0 | — | -0.61 | (0.20) | -0.67 | (0.22) |
| $\gamma_2$[Black] | 0 | — | -1.65 | (0.31) | -1.80 | (0.34) |
| $\gamma_3$[Black] × [Female] | 0 | — | 0.66 | (0.32) | 2.09 | (0.86) |
| $\psi$ | 2.47 | (0.84) | 1.88 | (0.59) | 2.27 | (0.74) |
| 对数似然值 | -2 002.76 | | -1 956.25 | | -1 954.89 | |
| 偏差 | 204.69 | | 111.68 | | 108.96 | |
| 皮尔逊 $X^2$ | 190.15 | | 102.69 | | 100.00 | |

根据 7.3.1 节中的经验贝叶斯法,我们可根据对这 4 个题器的回答来估计被测者的能力。对于 $M1$ 和 $M2$,对所有可能的回答模式的经验贝叶斯估计列在表 9.7 中。能力可以被理解为对于题器 1 回答正确的对数概率(因为 $\beta_1=0,\lambda_1=1$),0 表示 50% 的概率,±1 分别表示概率为 73% 和 27%。对于 $M2$,对于能力的估计依赖于分组,如由于黑人男性比白人男性能力均值低,因此,在同样的表现下,黑人男性的"得分"比白人男性低。从统计上说,这样的估计可能优于忽略协变量的情形,但是计分的方法显然是不公平的!

表 9.7 能力水平的经验贝叶斯估计

| 题器响应 | | | | M1 | M2 | | | |
|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 全部分组 | 白人男性 | 白人女性 | 黑人男性 | 黑人女性 |
| 0 | 0 | 0 | 0 | -1.18 | -0.53 | -0.85 | -1.43 | -1.41 |
| 0 | 0 | 0 | 1 | -0.13 | 0.27 | -0.02 | -0.53 | -0.51 |
| 0 | 0 | 1 | 0 | -0.34 | 0.20 | -0.09 | -0.60 | -0.58 |
| 0 | 0 | 1 | 1 | 0.62 | 0.98 | 0.70 | 0.21 | 0.23 |
| 0 | 1 | 0 | 0 | -0.40 | 0.11 | -0.18 | -0.71 | -0.68 |
| 0 | 1 | 0 | 1 | 0.57 | 0.88 | 0.60 | 0.11 | 0.13 |
| 0 | 1 | 1 | 0 | 0.36 | 0.82 | 0.53 | 0.04 | 0.07 |
| 0 | 1 | 1 | 1 | 1.31 | 1.62 | 1.32 | 0.82 | 0.84 |
| 1 | 0 | 0 | 0 | -0.05 | 0.38 | 0.09 | -0.42 | -0.39 |
| 1 | 0 | 0 | 1 | 0.90 | 1.16 | 0.87 | 0.38 | 0.40 |
| 1 | 0 | 1 | 0 | 0.69 | 1.09 | 0.80 | 0.31 | 0.34 |
| 1 | 0 | 1 | 1 | 1.66 | 1.91 | 1.60 | 1.09 | 1.12 |
| 1 | 1 | 0 | 0 | 0.64 | 0.99 | 0.71 | 0.22 | 0.24 |
| 1 | 1 | 0 | 1 | 1.61 | 1.80 | 1.50 | 0.99 | 1.02 |
| 1 | 1 | 1 | 0 | 1.39 | 1.73 | 1.43 | 0.93 | 0.95 |
| 1 | 1 | 1 | 1 | 2.45 | 2.64 | 2.29 | 1.73 | 1.76 |

## 9.5 尼古丁咀嚼胶与戒烟：元分析

### 9.5.1 介绍

在临床诊疗中，对各种诊疗法的效率及各种症状进行系统的综述是非常重要的。这种综述越来越普遍，并构成了“实证医学”（或证据医学）的非常重要的部分。萨科特等（Sackett et al.，1991）向临床医生提出了如下建议：

> 只有针对临床问题进行了严格的科学的综述之后，你才能尝试去回答（这个问题），你的时间才可能更好地花在研究上面，而非仅仅抓住某些（很可能是歪曲的）文献。

系统综述的重要性也体现在1993年科克伦协作组织（Cochrane Collaboration）的成立上，该组织针对医学研究的绝大部分领域建立并维护更新数量巨大的临床诊疗综述，除此之外还设立了指导方针，提供培训，等等。

元分析就是用统计的方法来合并来自多项研究中的医症，以获得对诊疗效应的总体估计。尽管现代元分析起源于教育学和心理学（如 Glass，1976；Hunt，1997），它最近在医学研究中的迅速发展在生物统计学中也带来了兴趣高潮。

本书我们讨论用尼古丁替代疗法进行戒烟的临床诊疗元分析，数据[1]来自希拉吉等（Silagy et al.，2003）。根据希拉吉等，我们对使用尼古丁咀嚼胶（而非尼古丁药帖），同时

1 数据可从 gllamm.org/books 下载。

配合高度支持的包括正式治疗或"评估及强化"拜访法的研究进行单独分析。

在每一项研究中,患者被随机地分配到试验组(给予尼古丁咀嚼胶)或控制组中。在大多数研究中,分配到控制组的患者收到与尼古丁咀嚼胶具有相同外观但不具有尼古丁成分的安慰剂。但在有些研究中,分配到控制组的患者没有收到咀嚼胶。在诊疗至少6个月后的戒烟效果是研究所关心的结果。在每项诊疗中,最严格的节制定义被采纳。这些诊疗的结果可以被归纳为2乘2表格,该表格也可根据表9.8的行数据推导出来。

我们将分析总体的优比估计,即试验组中戒烟的比率除以控制组中戒烟的比率。对于某项研究$j$,其优比估计为

$$o_j = \frac{d_{1j}/(n_{1j} - d_{1j})}{d_{0j}/(n_{0j} - d_{0j})} \tag{9.2}$$

其中,$d_{1j}$和$d_{0j}$分别是试验组和控制组中成功戒烟的个数,$n_{1j}$和$n_{0j}$分别为试验组和控制组的全部人数。关于试验效应的其他量度法还包括风险比(risk ratio)和风险差异(risk difference)。

**表9.8　随机化的尼古丁咀嚼胶和戒烟研究**

| 研究 | | 试验组 | | 控制组 | |
|---|---|---|---|---|---|
| | | 戒烟数 $d_1$ | 总数 $n_1$ | 戒烟数 $d_0$ | 总数 $n_0$ |
| Blondal | 1989 | 37 | 92 | 24 | 90 |
| Campbell | 1991 | 21 | 107 | 21 | 105 |
| Fagerstrom | 1982 | 30 | 50 | 23 | 50 |
| Fee | 1982 | 23 | 180 | 15 | 172 |
| Garcia | 1989 | 21 | 68 | 5 | 38 |
| Garvey | 2000 | 75 | 405 | 17 | 203 |
| Gross | 1995 | 37 | 131 | 6 | 46 |
| Hall | 1985 | 18 | 41 | 10 | 36 |
| Hall | 1987 | 30 | 71 | 14 | 68 |
| Hall | 1996 | 24 | 98 | 28 | 103 |
| Hjalmarson | 1984 | 31 | 106 | 16 | 100 |
| Huber | 1988 | 31 | 54 | 11 | 60 |
| Jarvis | 1982 | 22 | 58 | 9 | 58 |
| Jensen | 1991 | 90 | 211 | 28 | 82 |
| Killen | 1984 | 16 | 44 | 6 | 20 |
| Killen | 1990 | 129 | 600 | 112 | 617 |
| Malcolm | 1980 | 6 | 73 | 3 | 121 |
| McGovern | 1992 | 51 | 146 | 40 | 127 |
| Nakamura | 1990 | 13 | 30 | 5 | 30 |
| Niaura | 1994 | 5 | 84 | 4 | 89 |
| Niaura | 1999 | 1 | 31 | 2 | 31 |
| Pirie | 1992 | 75 | 206 | 50 | 211 |
| Puska | 1979 | 29 | 116 | 21 | 113 |
| Schneider | 1985 | 9 | 30 | 6 | 30 |
| Tonnesen | 1988 | 23 | 60 | 12 | 53 |
| Villa | 1999 | 11 | 21 | 10 | 26 |
| Zelman | 1992 | 23 | 58 | 18 | 58 |

资料来源:Silagy et al. (2003)。

### 9.5.2 元分析方法的视角

关于元分析存在两种本质上有所区别的视角:固定效应及随机效应。固定效应元分析假设存在一个真实的试验效应,并且任何研究所估计的试验效应之间的变差完全是由研究内的抽样变差所造成。在此请注意,"固定效应"这个词有点误导性,这是因为固定效应通常表示对于每个群组(在本例指每项研究)的效应是固定的,参见 3.6.1 节。对公共的试验效应的假设通常采用科克伦同质性 $Q$ 检验(Cochran $Q$-test of homogeneity)(如,Cochran,1950;DerSimonian and Laird, 1986)。

与此相反,随机效应元分析假设真实的试验效应在不同研究中是有变化的。这种变差可能是由于总体中的差异以及诊疗过程的因素,如药物用量、诊疗期长短、对结果的定义和量度及进一步跟踪的时间等。因此,元分析的目的就成了对一个想象的研究总体来估计平均试验效应。

弗雷斯(Fleiss,1993)和拜利(Bailey,1987)从两个方面来讨论如何从这两种相互竞争的视角中进行选择。首先,随机效应视角试图将结论推广到所有研究总体,包括未来的研究;而固定效应视角将结论限制在分析所使用的研究中。其次,随机效应视角允许研究项目之间存在变差,而固定效应视角则与之相反。在很多情形下,很多研究与其他同类研究具有根本上的不同,因此,将所有研究的效应混为一谈很容易被看做是荒谬的做法,从而不可能存在一个真实的试验效应。

在此,我们采纳随机效应视角,因为我们认为试验效应在研究中没有差异这个假设不符合先验知识。例如,"Blondal 1989"的研究使用了含 4 mg 尼古丁的咀嚼胶,为期 1 个月;"Barcia 1989"的研究则使用含 2 mg 尼古丁的咀嚼胶,为期 3 ~4 个月。前一项研究采用的是连续 12 个月的戒烟期,而后一项研究则采用连续 6 个月的戒烟期。这些研究在吸烟者所接受的支持的强度、吸烟者的类型上面也有所不同。例如,"Campbell 1991"的研究只针对由于吸烟导致疾病的患者,而其他大多数研究针对任何对戒烟感兴趣的吸烟者。更有甚者,有些研究在多个国家范围内进行,包括冰岛、瑞典、西班牙和美国。

无论是固定效应还是随机效应视角,在元分析中,其重点都在于从研究层次上估计试验效应,而非在原始的患者层次上。当所关心的效应尺度是优比时(如本例子),经常采用对数优比,而非优比,因为其样本分布更接近于正态分布。随机效应元分析包含对下述线性随机截距模型的估计:

$$\ln(o_j) = \beta_0 + \zeta_{0j} + \epsilon_j, \epsilon_j \sim \mathrm{N}(0, \theta_j) \tag{9.3}$$

其中,$o_j$ 正如式(9.2)的定义,为第 $j$ 项研究的优比估计值;$\beta_0$ 为关注的对数优比的均值;$\zeta_{0j}$为随机截距,表示该研究的真实对数优比与对数优比均值的偏差。研究内标准差$\sqrt{\theta_j}$被设为用沃尔夫法(Woolf,1955)估计的对数优比的标准误差,即

$$\sqrt{\theta_j} = \sqrt{\frac{1}{d_{0j}} + \frac{1}{n_{0j} - d_{0j}} + \frac{1}{d_{1j}} + \frac{1}{n_{1j} - d_{1j}}} \tag{9.4}$$

近年来,通过引入研究相关的协变量,如式(9.3)的药物用量,从而消除研究间异质性的"元回归"吸引了很多研究者的关注(如 Berkey et al., 1995; van Houwelingen et al., 2002)。

然而,由于正态性假设经常被违背,因而分析研究层次上试验效应的大小是有问题的。在对数优比的例子中,也可能出现只具有很小样本的研究或者感兴趣的因变量属于极小概率事件的情形。如果 $d_0$ 且/或 $d_1$ 为 0,那就需要特别的处理,如将计数值增加0.5(为了减少极小计数值的偏误所采用的一种处理办法,加尔特和兹维费尔(Gart and

Zweifel,1967)也作此建议)。从而,针对观测到的患者层次的响应直接建模显得更佳。但奇怪的是,对研究层次上的估计进行分析并不仅仅在应用类的论文中很普遍,在方法论研究中也是如此(如 Normand, 1999),包括贝叶斯诊断法(见 Carlin, 1992; DuMouchel et al. ,1996; Gelman et al. , 2003),在这个方法中用马尔可夫蒙特卡洛方法建立"合理"的模型其实是相当直接的。只有当患者层次的数据无法获取时,分析研究层次上效应的大小才有用武之地(另见 Chalmers,1993)。

### 9.5.3 患者层次数据的随机效应模型

对于患者层次的数据,我们定义 $i$ 表示患者,$j$ 表示研究,阿奎斯提和哈泽尔(Agresti and Hartzel, 2000)提出了下面这个随机系数模型:

$$\text{logit}(\Pr(y_{ij}=1 \mid x_{ij},\zeta_{0j},\zeta_{1j}) = \beta_0+\beta_1 x_{ij}+\zeta_{0j}+\zeta_{1j}x_{ij}$$

其中

$$x_{ij}=\begin{cases}0.5 & \text{对于试验组}\\ -0.5 & \text{对于控制组}\end{cases}$$

且

$$(\zeta_{0j},\zeta_{1j})' \sim N_2(\mathbf{0},\overline{\boldsymbol{\Psi}})$$

此处,$\beta_0$ 和 $\zeta_{0j}$ 分别为固定和随机截距,$\beta_1$ 和 $\zeta_{1j}$ 为 $x_{ij}$ 的固定斜率和随机斜率。$\beta_1$ 表示所感兴趣的对数优比,$\beta_1+\zeta_{1j}$ 表示第 $j$ 项研究的"真实"对数优比。与研究相关的截距有时候也被看做是固定的,如 Turner et al. (2000) 以及 Thompson et al. (2001)。

阿奎斯提和哈泽尔假定当 $\psi_{10}=0$ 时,随机截距与斜率不相关。在这个例子中,请注意 $x_{ij}$ 的编码是很重要的,如一个模型的 $x_{ij}=0,1$,与另一个 $x_{ij}=-0.5, 0.5$ 的模型是不等价的(见 3.2.2 节)。阿奎斯提和哈泽尔认为后一种编码方法(或称为"中心编码法")的优点在于当 $\psi_{10}=0$ 时,对数比率的总体方差 $\text{Var}(\zeta_{0j}+\zeta_{1j}x_{ij})$ 在所有组中都是相同的。

我们还需进一步检查该尼古丁咀嚼胶数据零相关假设的效度。这个模型是用 20 点自适应积分法进行的最大似然估计。随机截距和斜率之间的相关系数估计值为 0.17,相关模型与零相关模型在对数似然值上的差异仅有 0.05。因此,在表 9.9 中,我们在"经验贝叶斯"(最大似然)栏下列出 $\psi_{10}=0$ 模型的估计值。显然,有足够的证据说明尼古丁咀嚼胶有助于增加戒烟的比率,所估计的优比为 exp(0.57) = 1.77。但是,总的来看,戒烟发生率存在一些异质性,从所估计的 $\hat{\psi}_{00}^{1/2}=0.70$,以及试验效应仅有很小的变差(其估计值 $\hat{\psi}_{11}^{1/2}=0.22$)可以看出。

**表 9.9 经验贝叶斯、完全贝叶斯和 NPMLE 估计**

| 参 数 | 经验贝叶斯 | | 完全贝叶斯 | | NPMLE | |
|---|---|---|---|---|---|---|
| | Est | (SE) | Est | (SE) | Est | (SE) |
| 固定部分 | | | | | | |
| $\beta_0$[Cons] | -1.16 | (0.14) | -1.17 | (0.15) | -1.20 | (0.15) |
| $\beta_1$[Treat] | 0.57 | (0.09) | 0.59 | (0.09) | 0.59 | (0.10) |
| 随机部分 | | | | | | |
| $\sqrt{\psi_{00}}$[Cons] | 0.70 | (0.11) | 0.73 | (0.12) | — | |
| $\sqrt{\psi_{11}}$[Treat] | 0.22 | (0.10) | 0.20 | (0.10) | — | |

我们还可考虑用6.11节所描述的无信息先验完整贝叶斯法。$\beta_0$ 与 $\beta_1$ 的先验分布分别被设定为 $N(0,10^6)$，$\zeta_{0j}$与$\zeta_{1j}$的先验分布分别被设定为 $N(0,\psi_{11})$。$\psi_{00}$与$\psi_{11}$的假设被设定为IG(0.001,0.001)，其中IG为逆伽玛密度(见6.11.5节)。请注意，这种设定与BUGS手册(Spiegelhalter et al., 1996b)第一卷第十章中的元分析例子非常相像，区别在于我们将截距处理成带有先验及超先验方差的随机效应，而在手册中则将之处理为“固定效应”，不带有超先验。正如BUGS的实现机制，吉布斯抽样用于估计参数(见6.11.5节)。首先进行10 000次的迭代预处理，然后再进行1 000次迭代以计算出均值和标准差。计算结果显示在表9.9“完整贝叶斯”一栏中，并且与经验贝叶斯(最大似然估计)的计算结果非常接近。

我们可用式(9.2)与式(9.4)来计算每个研究的对数优比以及标准误。然而，如果我们相信贝叶斯随机效应模型的估计，那么所有关于单个对数优比 $\beta_1+\zeta_{1j}$的推导应当基于它们的边缘后验分布，并考虑整体模型的其他参数，在这个例子中，即$\beta_0$，$\psi_{00}$以及$\psi_{11}$。根据MCMC算法(见6.11.5节)，这等价于用对数优比 $\beta_1^{(r)}+\zeta_{1j}^{(r)}$样本的经验分布。在经验贝叶斯中，在其他参数与最大似然估计相等的情况下(见7.3.1节)，使用了$\zeta_{1j}$的条件后验分布。因此，所估计的对数优比就成为$\hat{\beta}_1+\tilde{\zeta}_{1j}$，其中，$\tilde{\zeta}_{1j}$为经验贝叶斯估计(条件后验的均值)。

贝叶斯学派使用可信区间(credible intervals)，而非置信区间。对于95%的可信区间，一个参数处于此区间的后验概率为95%。图9.6显示了每项研究$\beta_1+\zeta_{1j}$的真实效应的近似95%贝叶斯可信区间(基于1 000次抽取所估计的后验分布)，及其经验贝叶斯可信区间。与完整贝叶斯区间不同，经验贝叶斯区间是通过将模型参数设为已知，并且假定随机斜率的后验分布为正态而推导出。在此，区间被定义为后验均值 ± 1.96 与后验标准差的乘积。由于完整贝叶斯考虑了参数的不确定性，其可信区间本应更宽一些。但是，由于“Huber 1988”的存在，这些差别变得非常微小。显示在 $x$ 轴上的原始对数优比(raw log odds ratio)$\ln(o_j)$比经验贝叶斯以及完整贝叶斯更加偏离一般的对数比率(显示为垂直实线)。这是由于7.3.1节所讨论的收缩所造成的。

通常假定随机截距与斜率为二元正态分布，但是艾特金(Aitkin, 1999b)采用非参数最大似然估计法(NPMLE)(见4.4.2节)，从而可以无须假定随机截距与斜率的联合分布形态。在此，对于混合回归模型，使用了一个离散分布，其原点为$\zeta_{0j}=e_{oc}$，$\zeta_{1j}=e_{1c}$，质量或概率为$\pi_c$，$c=1,\cdots,C$。质量$C$的个数确定似然值的最大化。请注意，截距与斜率就不再是互相独立的了。使用6.5节的加托导数法，使得$C$每次递增1直到导数为负数，我们得到$C=10$。

如表9.9所示，NPMLE一栏下的截距与斜率均值估计与使用二元正态分布假设以及$\psi_{10}=0$条件下的估计结果非常相似。NPMLE的对数似然估计值为 -3 061.5，而二元正态分布假设以及$\psi_{10}=0$条件下的对数似然估计值为 -3 074.2。在NPMLE法中，随机效应的方差以及协方差并不是模型的参数，但是可以从离散分布中推导出来。随机截距的标准差为0.76，随机斜率的标准差为0.30，相关系数为0.08。

如图9.7所示为NPMLE质量。在上图中，圆圈的位置为$\hat{e}_{0c}$和$\hat{e}_{1c}$，其区域大小与概率$\hat{\pi}_c$成比例。而在下图中，概率用长钉的长度来表示。如图9.8所示为控制组与试验组(咀嚼胶)的对数比率。线条的粗细表示概率的大小，其数字为概率百分比。

综上所述，3种方法得到的估计结果几乎一致。就总体结论来说，使用尼古丁咀嚼胶使得戒烟的比率增加80%。3种方法所得到的优比的置信区间以及可信区间(根据估计

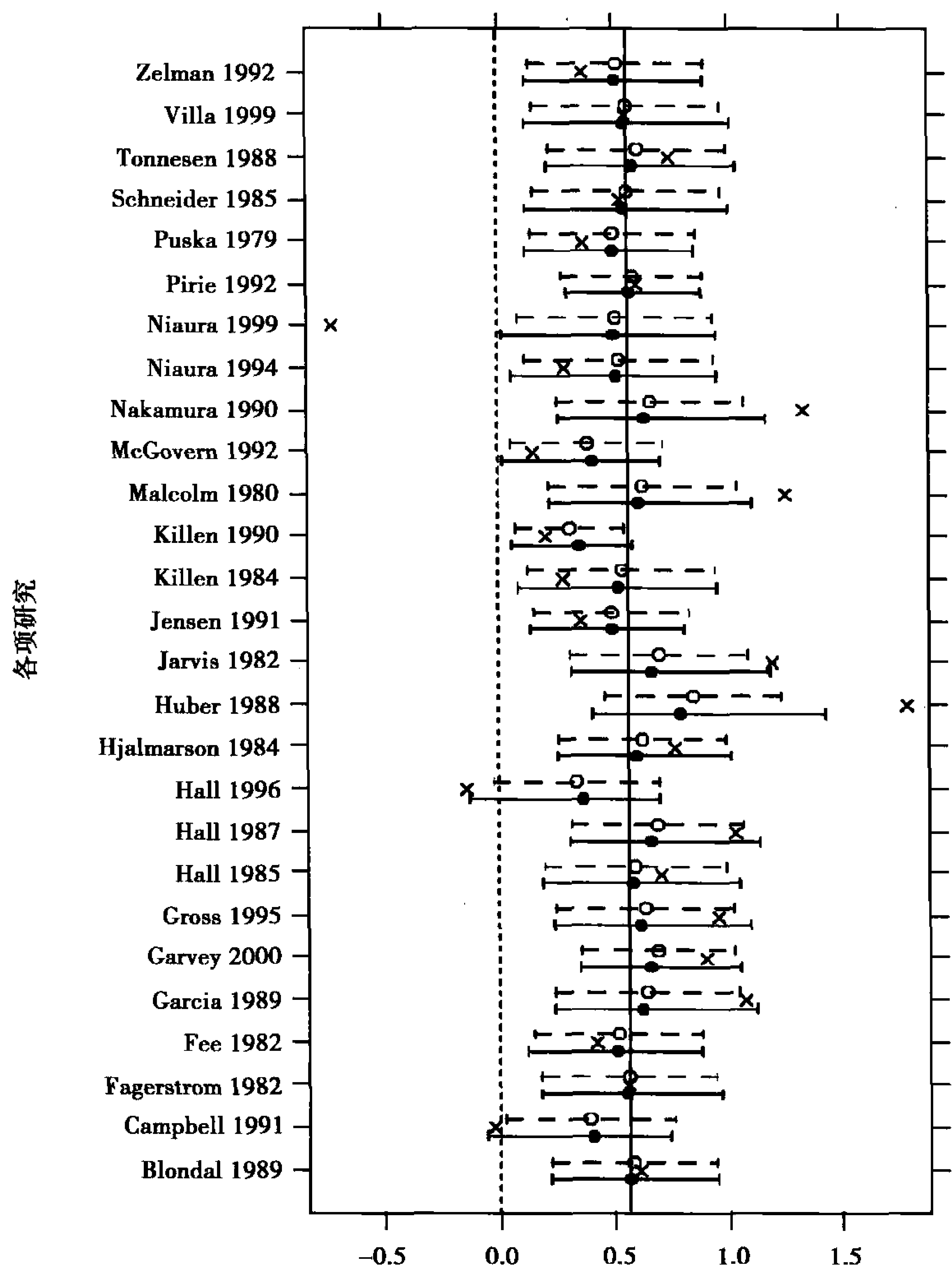



**图 9.6 各项研究的真实效应 $\beta_1+\zeta_{1j}$ 的经验贝叶斯(虚线)及完整贝叶斯(实线)区间。$x$ 轴为原始对数优比 $\ln(O_j)$,垂直实线轴为对数优比均值 $\hat{\beta}_1$ 的最大似然估计,在 $O$ 点位置的垂直虚线表示没有试验效应**

的对数优比及其标准误推导而出)几乎相同,经验贝叶斯的区间为 1.5 至 2.1。在各项研究中,从试验效应上看不到太多的异质性。

必须明确的非常重要的一点是,针对元分析方法还存在很多批评,请见 Thompson and Pocock (1991) 以及 Oakes (1993)。其中,一个广为承认的问题是发表偏差(publication bias)。小型研究如果没有特殊发现则通常难于发表,从而导致在元分析中过高估计试验效应(如 Sterlin, 1959; Sutton et al., 2000)。

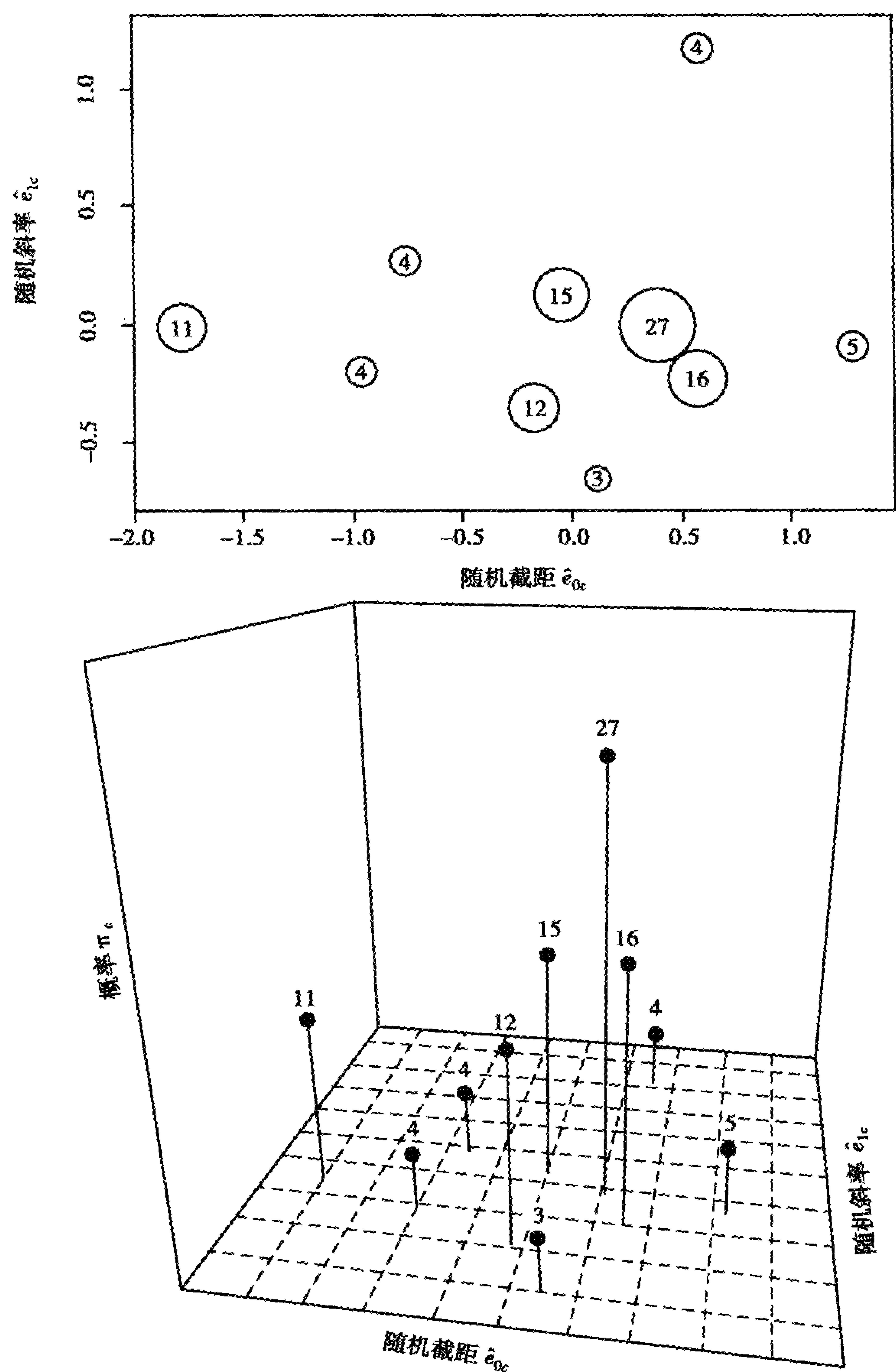



图 9.7 NPML 质量(所显示的数字为概率百分比)

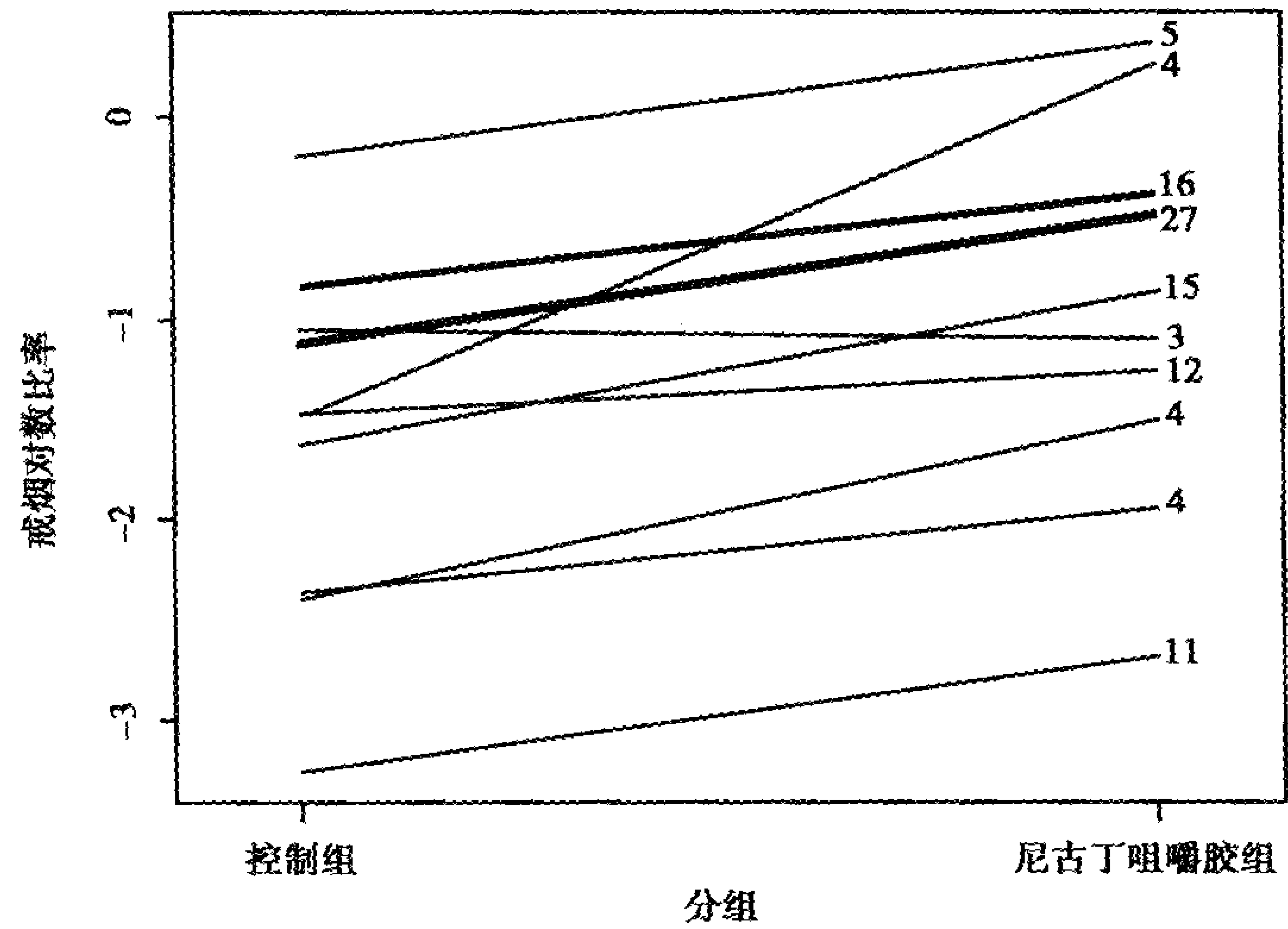



图 9.8 对数比率估计值

## 9.6 妻子的就业转换:马尔可夫模型及未观测到的异质性

我们将要分析一个妻子的工作数据[1,2],该数据来自一项追踪性研究或重复测量研究。戴维斯等(Davies et al., 1992)以及戴维斯(Davies, 1993)的"社会变迁与经济生活动力"研究,记录了家庭中妻子从结婚至调查之时(1987 年)每个月的工作状况。我们此处使用罗奇代尔(Rochdale)的 6 城市数据子集。

响应变量为妻子的工作状态,在职(1)或未在职(0)。解释变量都是随时点变化的变量:

- [HUnemp] 虚拟变量,丈夫的工作状态,1 为未在职,0 为在职。
- [Time] 从研究开始观测的月数。
- [Child1] 是否有 1 岁以下的子女,虚拟变量。
- [Child5] 是否有 5 岁以下的子女,虚拟变量。
- [Age] 妻子的年龄。
- [Agesq] 妻子年龄的平方。

人们在过去所经历的事件很可能在将来继续经历,关于这个规律存在两个解释。其中一个解释是"因果论":工作经历改变人们,曾经有过的工作经历会对妻子将来的工作概率发生影响,产生时间上的依赖性。赫克曼称之为"**真实状态依赖**(true state dependence)"(Heckman, 1978, 1981ac)。这种因果解释比较流行,但有时过于肤浅,而备择种解释同样似是而非。在此例中,观测到的时间依赖性来自于表示未观测到的异质性的那种不变的因素。也就是说,更高的就业概率很可能是由于未被观测到的妻子们的某些特性,有些人不管其以前的工作经历如何,本身就具有更高的未来就业概率。这被赫克曼称为"**虚假状态依赖**(spurious state dependence)",他认为在观测性研究中区分真实状态依赖与虚假状态依赖是非常重要的。

首先考虑一下传统的一阶离散时间马尔可夫模型,给定前一个就业状态 $y_{t-1,j}$,在时间 $t$ 的就业状态 $y_{tj}$条件依赖于就业历史及当前的协变量。使用对数概率链接函数,该模型可表示为

$$\Pr(y_{tj}=1\mid \mathbf{x}_{tj},y_{t-1,j})=\frac{\exp(\mathbf{x}'_{tj}\boldsymbol{\beta}+\gamma y_{t-1,j})}{1+\exp(\mathbf{x}'_{tj}\boldsymbol{\beta}+\gamma y_{t-1,j})}$$

通过使用随机截距 $\zeta_j$,可将未观测到的异质性包含到此马尔可夫模型中(见 Heckman, 1981a)

$$\Pr(y_{tj}=1\mid \mathbf{x}_{tj},y_{t-1,j},\zeta_j)=\frac{\exp(\mathbf{x}'_{tj}\boldsymbol{\beta}+\gamma y_{t-1,j}+\zeta_j)}{1+\exp(\mathbf{x}'_{tj}\boldsymbol{\beta}+\gamma y_{t-1,j}+\zeta_j)}$$

根据该模型,当 $\gamma\neq 0$,则存在真实状态依赖;而当 $\gamma=0$,则存在虚假状态依赖。

通过指定一个因子模型,我们可允许随机截距方差依赖于前一个状态,即

$$\Pr(y_{tj}=1\mid \mathbf{x}_{it},y_{t-1,j}=0,\zeta_j)=\frac{\exp(\mathbf{x}'_{tj}\boldsymbol{\beta}+\zeta_j)}{1+\exp(\mathbf{x}'_{tj}\boldsymbol{\beta}+\zeta_j)}$$

与

---

1 感谢戴夫·斯托特(Dave Stott)提供数据。

2 数据可从 gllamm.org/books 下载。

$$\Pr(y_{tj}=1\mid \mathbf{x}_{tj},y_{t-1,j}=1,\zeta_j)=\frac{\exp(\mathbf{x}'_{tj}\boldsymbol{\beta}+\gamma+\lambda\zeta_j)}{1+\exp(\mathbf{x}'_{tj}\boldsymbol{\beta}+\lambda\zeta_j)}$$

第1个公式的概率条件依赖于对象之前处于状态0,而第2个公式的概率条件依赖于对象处于状态1。其他两个转换概率为 $\Pr(y_{tj}=0\mid y_{t-1,j}=0)=1-\Pr(y_{tj}=1\mid y_{t-1,j}=0)$ 以及 $\Pr(y_{tj}=0\mid y_{t-1,j}=1)=1-\Pr(y_{tj}=1\mid y_{t-1,j}=1)$。因子 $\zeta_j$ 表示在转换过程中对象自身的未观测到的异质性,当 $y_{t-1,j}=1$ 时,方差为 $\lambda^2\psi$;当 $y_{t-1,j}=0$ 时,方差为 $\psi$。

这3个模型分别用 $M1$,$M2$,$M3$ 来表示,其参数估计显示在表9.10中。效应估计值最大的为[Child1]和[HUnemp],它们均降低妻子就业的比率。1岁以内的子女对妻子的就业存在负效应,这毫不令人奇怪。而[HUnemp]的负效应很可能是由于妻子就业后会显著降低丈夫的失业补偿或者伤害男性的自尊。比较模型1和模型3,有明显的证据说明存在状态依赖及未观测到的异质性,当之前的就业状态为未在职时,随机效应方差显著增大。

**表9.10 包含及未包含未观测到的异质性的简单马尔可夫模型估计**

| 参 数 | $M1$ | | $M2$ | | $M3$ | |
|---|---|---|---|---|---|---|
| | Est | (SE) | Est | (SE) | Est | (SE) |
| 固定部分 | | | | | | |
| $\beta_0$[Cons] | -1.277 | (0.240) | -1.114 | (0.299) | -1.222 | (0.339) |
| $\beta_1$[HUnemp] | -1.406 | (0.373) | -1.512 | (0.412) | -1.524 | (0.388) |
| $\beta_2$[Time] | -0.012 | (0.025) | -0.011 | (0.026) | 0.006 | (0.032) |
| $\beta_3$[Child1] | -3.008 | (0.392) | -2.953 | (0.408) | -2.767 | (0.409) |
| $\beta_4$[Child5] | -0.165 | (0.253) | -0.241 | (0.272) | -0.383 | (0.294) |
| $\beta_5$[Age] | -0.005 | (0.014) | 0.000 | (0.016) | -0.009 | (0.018) |
| $\beta_6$[Agesq] | -0.001 | (0.001) | -0.001 | (0.001) | -0.002 | (0.001) |
| $\gamma$[Lag] | 4.391 | (0.209) | 4.226 | (0.264) | 4.380 | (0.326) |
| 随机部分 | | | | | | |
| $\psi$ | — | — | 0.308 | (0.326) | 2.177 | (0.932) |
| $\lambda$ | — | — | — | — | -0.119 | (0.164) |
| 对数似然值 | -411.50 | | -410.89 | | -401.25 | |

由弗朗西斯等(Francis et al.,1996)提出的模型 $M4$ 允许回归参数以及未观测到的异质性的效应依赖于前一个状态,即

$$\Pr(y_{tj}=1\mid \mathbf{x}_{tj},y_{t-1,j}=0,\zeta_j)=\frac{\exp(\mathbf{x}'_{tj}\boldsymbol{\beta}^0+\zeta_j)}{1+\exp(\mathbf{x}'_{tj}\boldsymbol{\beta}^0+\zeta_j)}$$

与

$$\Pr(y_{tj}=1\mid \mathbf{x}_{tj},y_{t-1,j}=1,\zeta_j)=\frac{\exp(\mathbf{x}'_{tj}\boldsymbol{\beta}^1+\lambda\zeta_j)}{1+\exp(\mathbf{x}'_{tj}\boldsymbol{\beta}^1+\lambda\zeta_j)}$$

当前一个状态为0时,协变量 $\boldsymbol{x}_{tj}$ 的固定效应为 $\boldsymbol{\beta}^0$;当前一个状态为1时,则为 $\boldsymbol{\beta}^1$。在消除了未观测到的异质性之后,如果 $\boldsymbol{\beta}^0\neq\boldsymbol{\beta}^1$,则存在真实状态依赖。在这个模型中,如果在引入未观测到的异质性之前 $\boldsymbol{\beta}^0\neq\boldsymbol{\beta}^1$,且在引入未观测到的异质性之后 $\boldsymbol{\beta}^0=\boldsymbol{\beta}^1$,则存在虚假状态依赖。根据4.2.3节的GRC方程式,该模型可重写为

$$\operatorname{logit}[\Pr(y_{tj}=1 \mid \mathbf{x}_{tj}, y_{t-1,j}, \zeta_j)] = \nu_{tj}$$

其中,线性估计值为

$$\nu_{tj} = (1-y_{t-1,j})\mathbf{x}'_{tj}\boldsymbol{\beta}^0 + y_{t-1,j}\mathbf{x}'_{tj}\boldsymbol{\beta}^1 + \zeta_j[(1-y_{t-1,j}) + \lambda y_{t-1,j}]$$

该一般马尔可夫模型的参数估计如表 9.11 所示。系数上最大的差异在[Child1]变量上,对于在职的妻子,[Child1] 明显增加了去职的比率(优比 exp(2.776) = 16.1);而对于未在职的妻子,则增加了其继续保持未在职状态的比率(优比 exp(1.439) =4.2)。

**表 9.11 一般马尔可夫模型估计 *M*4**

| 参 数 | 获得工作 $y_{t-1,j}=0, a=0$ | | 保持工作 $y_{t-1,j}=1, a=1$ | |
|---|---|---|---|---|
| | Est | (SE) | Est | (SE) |
| 固定部分 | | | | |
| $\beta_0^a$[Cons] | -1.521 | (0.429) | 3.391 | (0.353) |
| $\beta_1^a$[HUnemp] | -1.916 | (0.735) | -1.361 | (0.493) |
| $\beta_2^a$[Time] | 0.013 | (0.042) | -0.013 | (0.040) |
| $\beta_3^a$[Child1] | -1.439 | (0.695) | -2.776 | (0.490) |
| $\beta_4^a$[Child5] | -0.253 | (0.396) | -0.554 | (0.413) |
| $\beta_5^a$[Age] | -0.031 | (0.027) | 0.031 | (0.019) |
| $\beta_6^a$[Agesq] | 0.009 | (0.002) | -0.002 | (0.002) |
| 随机部分 | | | | |
| $\lambda$ | | | -0.212 | (0.223) |
| $\psi$ | | 1.411(0.717) | | |
| 对数似然值 | | -396.99 | | |

为了更好地解释这个依赖结构,我们将此模型更改为潜响应模型的表达方式,即

$$y_{tj}^* = \mathbf{x}'_{tj}\boldsymbol{\beta}^0 + \zeta_j + \epsilon_{tj} \quad 当 \quad y_{t-1,j}=0$$

与

$$y_{tj}^* = \mathbf{x}'_{tj}\boldsymbol{\beta}^1 + \lambda\zeta_j + \epsilon_{tj} \quad 当 \quad y_{t-1,j}=1$$

其中

$$y_{tj} = \begin{cases} 1 & 当\ y_{tj}^* > 0 \\ 0 & 其他 \end{cases}$$

很明显,任意两个不邻近的事件($s$ 与 $u$, $|s-u|>1$)的潜响应的相关系数为

$$\rho_{00} = \operatorname{Cor}(y_{sj}^*, y_{uj}^* \mid y_{s-1,j}=0, y_{u-1,j}=0, \mathbf{x}_{sj}, \mathbf{x}_{uj}) = \frac{\psi}{\psi+\pi^2/3}$$

$$\rho_{11} = \operatorname{Cor}(y_{sj}^*, y_{uj}^* \mid y_{s-1,j}=1, y_{u-1,j}=1, \mathbf{x}_{sj}, \mathbf{x}_{uj}) = \frac{\lambda^2\psi}{\lambda^2\psi+\pi^2/3}$$

且

$$\rho_{01} = \operatorname{Cor}(y_{sj}^*, y_{uj}^* \mid y_{s-1,j}=0, y_{u-1,j}=1, \mathbf{x}_{sj}, \mathbf{x}_{uj}) = \frac{\lambda\psi}{\sqrt{\lambda^2\psi+\pi^2/3}\sqrt{\psi+\pi^2/3}}$$

其中,$\rho_{00}$表示妻子从未在职转换为在职的倾向性的残相关系数,其估计值为 $\hat{\rho}_{00}=0.30$。$\rho_{11}$表示妻子继续保持在职的倾向性的相关系数,其估计值为 $\hat{\rho}_{11}=0.02$。$\rho_{01}$表示从未在职转变为在职的倾向性与已在职且继续保持在职的倾向性之间的相关系数,其估计值为 $\hat{\rho}_{01}=-0.08$。

在动态模型中,需要设置初始条件。初始条件通常包括外生的初始状态(为了简单起见,本例子采用此方法),或者处于平衡状态的过程。然而,这两者共同的问题都在于,所研究的过程通常并不是从其原初的状态开始,即所观测到的第 1 个状态并非是外生的。赫克曼(Heckman, 1981b)针对二项动态模型(dichotomous dynamic models)提出了一个特别的处理方法使之趋近于初始条件。关于此方法以及其他更多方法的讨论请见萧政(Hsiao, 2002)。

## 9.7 白靴兔计数:捕获-再捕获模型及异质性

为了确定一个总体的规模,经常采用捕获-再捕获方法。例如,为了计算一个城市中的海洛因吸食者,或者是在某个地理区域内的某些生物种群。捕获-再捕获方法的思路在于分不同次数或不同方法从一个总体中"捕获"个体,并标识被捕获的个体(对于动物种群,通常要求对所捕获的动物进行标记)。通过研究那些被捕获过至少一次的个体的被捕获历史,可以估计出从未被捕获的个体的数量,进而可估算总体的规模。该方法的假设前提是,在研究过程中总体规模恒定或者保持不变。

首先来看一个简单的两次捕获例子。如果所有的个体都有相同的被捕获的概率,且两次被捕获的概率独立,那么如果两次都被捕获的个体比例很高,就说明其总体规模很小。或者,也可以解释为,某些个体比其他个体更容易被捕获,而这些正是两次都被捕获的个体。仅仅通过两次捕获,我们无法从经验上区分上述这两种解释。通过进行更多次数的捕获,我们可以通过观测到的捕获历史来估计可捕获性的未观测到的异质性程度。如果不考虑未观测到的异质性,在对总体规模进行估计时会导致偏误(如 Otis et al., 1978),尽管可以通过研究设计或者将观测到的协变量包含到模型中来以降低偏误。

伯恩汉姆(Burnham)和库西瓦(Cuschwa )(见 Otis et al., 1978)为了估计阿拉斯加黑杉林中的白靴兔种群规模,设置了一系列捕捉陷阱。他们以 200 英尺[1]为间距,设置了 10×10 个陷阱。在连续 9 天的捕捉研究中,前 3 天并未抓获任何白靴兔。在后 6 天的捕捉中,共获得 68 个捕获-再捕获数据。如表 9.12 所示为在每个可能的捕获历史中,兔子被捕获的记录(1 = 捕获,0 = 未捕获)[2]。在表示 6 次均未捕获的单元格中,这个数字是空白的,我们的目的就是要估计这个数字。

表 9.12 白靴兔捕获-再捕获估计结果

| 捕获 6,5,4 | 捕获 3, 2, 1 | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 000 | 001 | 010 | 011 | 100 | 101 | 110 | 111 |
| 000 | ? | 3 | 6 | 0 | 5 | 1 | 0 | 0 |
| 001 | 3 | 2 | 3 | 0 | 0 | 1 | 0 | 0 |
| 010 | 4 | 2 | 3 | 1 | 0 | 1 | 0 | 0 |
| 011 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 100 | 4 | 1 | 1 | 1 | 2 | 0 | 2 | 0 |
| 101 | 4 | 0 | 3 | 0 | 1 | 0 | 2 | 0 |
| 110 | 2 | 0 | 1 | 0 | 1 | 0 | 1 | 0 |
| 111 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 2 |

资料来源:Agresti(1994)。

1 1 英尺 = 0.304 8 m。——编者注

2 数据可从 gllamm.org/books 下载。

估计总体规模的一个常用的方法是萨纳特南(Sanathanan, 1972)的条件法。首先,对于捕获历史 **y** 设定一个概率为 $\pi_y$ 的模型,**y** 代表每次捕获情况的一系列指标(0 或 1)。对于至少被捕获一次的个体,该模型的参数通过使观测到的捕获历史的条件似然(conditional likelihood)最大化来估计得到。对于至少被捕获一次的个体,其捕获历史的条件概率 **y** 为

$$\pi_{cy} = \pi_y/(1 - \pi_{0\cdots 0})$$

其中,$\pi_{0\cdots0}$为从未被捕获的概率。条件对数似然值则为

$$L_c = \sum_y n_y \ln \pi_{cy} = \sum_y n_y \ln \pi_y - \sum_y n_y \ln(1 - \pi_{0\cdots 0})$$

其中,加总是对所有观测到的捕获历史,即所有可能的捕获历史,除了 0⋯0,$n_y$ 为第 $y$ 次捕获的数量。对总体规模的估计就通过使至少被捕获一次的第 $n = \sum_y n_y$ 个个体的二项概率最大化来得到,其中成功的概率为$(1 - \pi_{0\cdots0})$,并且

$$\hat{N} = \frac{n}{1 - \pi_{0\cdots 0}}$$

在这个白靴兔捕获-再捕获研究中,很可能存在未观测到的异质性,原因有很多方面。我们可能认为异质性来自于行为的差异,如某些陷阱更有诱惑力,或者某些陷阱更容易被避开。更重要的是,拥有更广范围的草料场的兔子比那些仅有小范围草料场的兔子具有更高的掉入陷阱的可能性;生活在陷阱区域范围内的兔子逃避捕捉的可能性也更低。库尔和阿奎斯提(Coull and Agresti, 1999)使用随机截距对数概率模型来处理未观测到的异质性。在第 $i$ 次捕获中,个体 $j$ 被捕获的条件概率为

$$\text{logit}[\Pr(y_{ij} = 1 \mid \zeta_j)] = \beta_i + \zeta_j$$

其中,$\zeta_j$ 表示动物 $j$ 的"可捕获性",服从均值为 0,方差为 $\psi$ 的正态分布。这也就是3.3.4节中所讨论的单参数对数概率题器响应模型。对于捕获历史 $\pi_y$,以及 $y = (y_1, \cdots, y_I)'$,其概率为

$$\pi_y = \int \prod_{i=1}^{I} \frac{\exp(y_i(\beta_i + \sqrt{\psi} z))}{1 + \exp(\beta_i + \sqrt{\psi} z)} \phi(z) \mathrm{d}z$$

我们并不对随机截距做出正态分布的假设,相反,我们假设总体包含几个潜在的子类,每个子类具有不同的可捕获性程度,即允许 $\zeta_j$ 是离散的,使得

$$\pi_y = \sum_c \prod_{i=1}^{I} \frac{\exp(y_i(\beta_i + e_c))}{1 + \exp(\beta_i + e_c)} \pi_c$$

其中,$e_c$ 为子类 $c$ 的可捕获性,$\pi_c$ 为个体属于子类 $c$ 的概率。为了模型辨识的需要,我们将 $\zeta_j$ 的均值约束为 0,即

$$E(\zeta_j) = \sum_c \pi_c e_c = 0$$

使得方差 $\eta_j$ 为

$$\text{Var}(\zeta_j) = \sum_c \pi_c e_c^2$$

对于同质性总体模型,正态随机截距模型及二类模型的估计值如表 9.13 所示。二类模型的估计值为 $\hat{e}_1 = 4.14$, $\hat{e}_2 = -0.13$, $\hat{\pi}_1 = 0.03$, $\hat{\pi}_2 = 0.97$。将子类数目增加到 3 个,模型的条件对数似然值仅增加 0.30。请注意,这两个关于未观测到的异质性的模型所估计的总体规模有较大的差异(92 与 77),潜类方法的估计值更加接近传统的估计方法,尽管这两个异质性模型的拟合度非常接近。

模型假设响应之间的依赖完全是由于未观测到的异质性。然而,同样有理由假设前

一次的捕获直接影响下一次的捕获(状态依赖),特别是如果都用相同的捕获方法时。因此,哈金斯 (Huggins,1989)在随机截距模型中包含了一个关于前一次捕获 $x_{ij}$且随时间变化的指标([Prev]),如果动物在前次被捕获则 $x_{ij}=1$,否则 $x_{ij}=0$。

$$\text{logit}[\Pr(y_{ij}=1\mid\zeta_j,x_{ij})]=\beta_i+\zeta_j+\gamma x_{ij}$$

对于这个白靴兔数据以及正态分布随机截距 $\zeta_j$,模型的估计值如表 9.13 所示的最后一栏。$\hat{\gamma}=-1.10$,表示如果动物上次被捕获,则再次被捕获的概率降低。这个效应在 5% 的水平下($p=0.07$)并不是十分显著。该模型所估计的总体规模为 75,与用传统估计方法得到的结果相同。

**表 9.13　白靴兔总体的捕获-再捕获估计**

| 参　数 | 同质性总体 | | 随机截距 | | 二类模型 | | 随机截距及捕获历史 | |
|---|---|---|---|---|---|---|---|---|
| | Est | (SE) | Est | (SE) | Est | (SE) | Est | (SE) |
| $\beta_1$[Cap1] | -1.30 | (0.29) | -1.83 | (0.47) | -1.36 | (0.33) | -1.56 | (0.38) |
| $\beta_2$[Cap2] | -0.52 | (0.25) | -0.98 | (0.43) | -0.51 | (0.28) | -0.39 | (0.40) |
| $\beta_3$[Cap3] | -1.01 | (0.27) | -1.51 | (0.45) | -1.03 | (0.30) | -0.59 | (0.53) |
| $\beta_4$[Cap4] | -0.64 | (0.25) | -1.11 | (0.44) | -0.64 | (0.29) | -0.03 | (0.60) |
| $\beta_5$[Cap5] | -0.82 | (0.26) | -1.30 | (0.44) | -0.83 | (0.29) | -0.10 | (0.67) |
| $\beta_6$[Cap6] | -0.29 | (0.25) | -0.74 | (0.43) | -0.29 | (0.28) | 0.59 | (0.74) |
| $\gamma$[Prev] | — | — | — | — | — | — | -1.10 | (0.60) |
| Var($\zeta_j$) | — | — | 0.93 | (0.63) | 0.56 | — | 1.01 | (0.56) |
| $L_c$ | -254.5 | | -250.74 | | -249.16 | | -249.41 | |
| $\hat{N}$ | 75.1 | | 92.1 | | 77.1 | | 75.0 | |
| 95%CI | (69,84) | | (74,153) | | (70,88) | | (68,116) | |

科马克(Cormack,1992)提出用(非条件的)剖面似然来计算真实的总体规模 $N$ 的置信区间。该方法将 $n_{0\cdots0}$替换为不同的数值,通过最大化非条件似然来估计模型参数,并计算模型的偏差。置信限就为 $N=n+n_{0\cdots0}$的数值与用预先指定的数所产生的最小值之间的差。在这个例子中,在95%置信区间下该值为3.84。由于饱和模型的对数似然值以 $n_{0\cdots0}$变动,因此,要特别注意使用偏差而非对数似然。科马克证明,当 $n_{0\cdots0}$等于条件估计时,条件似然的参数估计及偏差与非条件似然的参数估计及偏差是相等的。

这 4 个模型的剖面偏差如图 9.9 所示。水平线表示最小偏差加 3.84。垂直线说明总体规模(整数 $N$)的近似 95% 置信限与偏差尽可能的接近,并不小于水平线显示的数值。根据表 9.13,置信区间比较宽,特别是对随机截距模型。

根据这些发现,我们可以得出一个比较保守的结论:白靴兔总体的规模为 68 ~ 153。

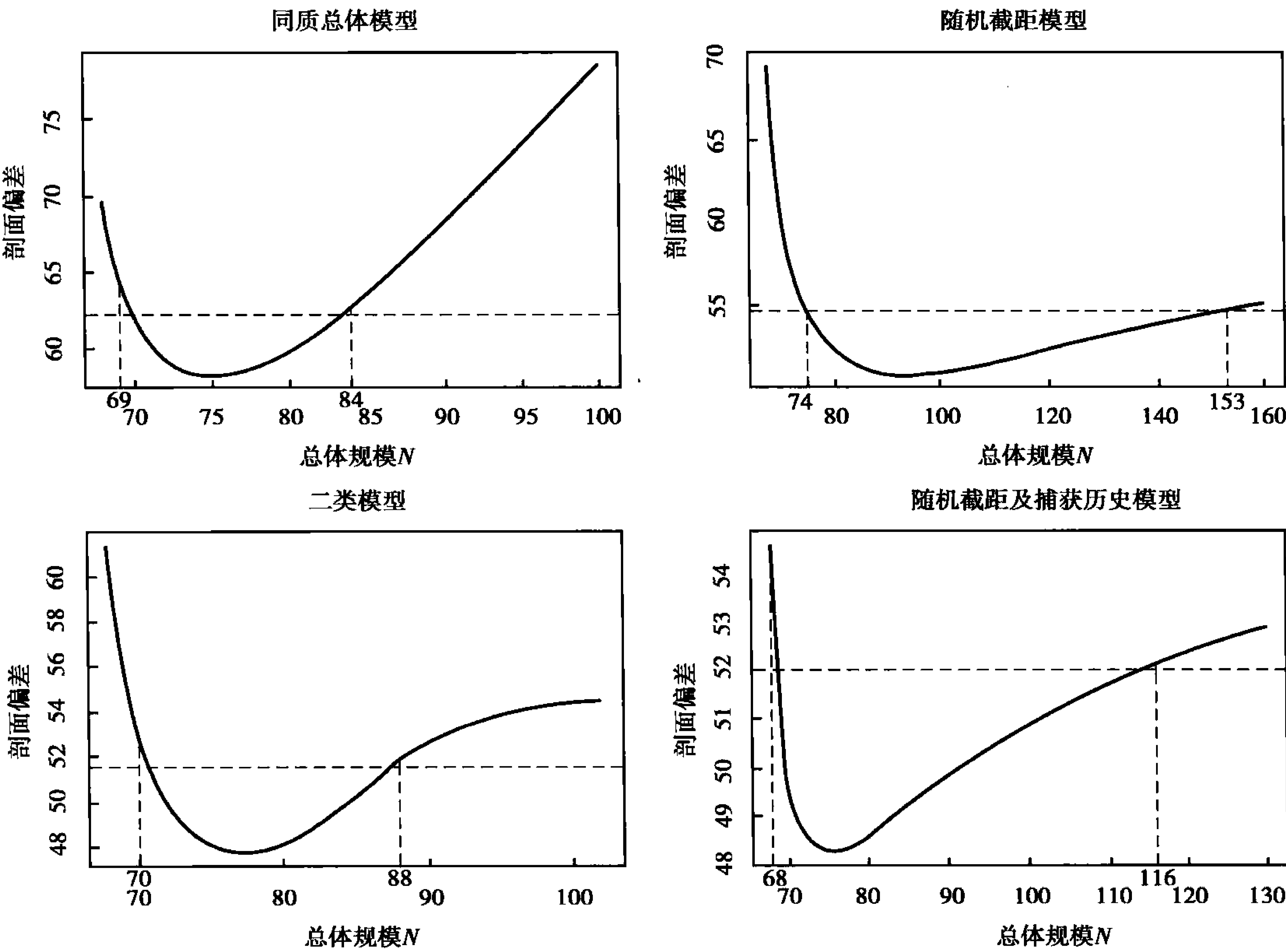



图 9.9 捕获-再捕获模型总体规模剖面偏差

## 9.8 对堕胎的态度:多层次题器响应模型

在英国社会态度跟踪抽样调查 1983—1986(社会及社区规划研究,1987)的数据[1]中,受访者被询问到,在下列情形中法律是否应该允许堕胎:

(1)[Woman] 该妇女自己决定不想要这个孩子。

(2)[Couple] 夫妇俩都同意不要孩子。

(3)[Marriage] 该妇女未婚且不愿与孩子的父亲结婚。

(4)[Financial] 该夫妇承担不起养育更多的孩子。

(5)[Defect] 婴儿可能不健全。

(6)[Risk] 妊娠会严重危害孕妇的健康。

(7)[Rape] 该妇女因被强奸而怀孕。

这个数据具有三层结构,调查的时点或“面板波次(panel waves)”嵌套于个体,再嵌套于调查区域。对于上述 7 个问题总共有 14 143 个回答,来自 4 个波次调查的 57 个区域的 734 个受访者。可以看出,这个多层次设计是非常不均衡的,因为 49% 的受访者在全部 4 个波次中回答了至少一个问题,12% 的受访者在 3 个波次中回答了至少一个,13% 在两个波次中回答了至少一个,25% 只在一个波次中回答。因而,个体无回答(unit nonresponse)非常普遍,但是如果访谈能够进行,题器则无回答(item nonresponse)发生的比例只占 7%。我们在此并不考虑个体无回答或题器无回答的因素,因而假设此数据为

1 数据来自英国数据中心。数据收集者及该中心不对此分析负责。

随机性缺失(MAR)。

前人对此数据的多层次分析使用原始加总分,或者使用根据题器响应模型所得到的响应得分(Knott et al., 1990; Wiggins et al., 1990)。然而,这种用构建的得分或者潜变量作为中介的方法已经被证明为存在问题,它将导致有偏标准误,并且经常产生不一致的参数估计(Skrondal and Laake, 2001)。因此,我们使用多层次因子模型并且对于二项变量使用对数概率链接函数(Rabe-Hesketh et al., 2004a)。模型在偏差上的变化将作为我们选择模型的依据。为了保证结果的可靠性,每个模型都用自适应积分法拟合多次,并与在每个维度上采用不同数目的计值点所得到的解进行比较。

首先,我们将重点放在研究主体间的异质性上面,然后再考虑不同调查区域间的异质性。人们除了持有所有题器都测量的"一般态度"因子 $\eta_{Gjk}^{(3)}$ 以外,还可能持有赞同堕胎的"极端情形"因子 $\eta_{Ejk}^{(3)}$,如婴儿很可能不健全,对孕妇的高度风险,或者因强奸而怀孕(题器5、题器6及题器7)。使用下标 $i$ 表示题器或假设的情形(第1层),$t$ 表示波次(第2层),$j$ 表示受访者(第3层),以及 $k$ 表示调查区域(第3层),双因子模型用如下GRC方程式:

$$\nu_{itjk} = \mathbf{d}_i'\boldsymbol{\beta} + \eta_{Gjk}^{(3)}\mathbf{d}_i'\boldsymbol{\lambda}_G + \eta_{Ejk}^{(3)}\boldsymbol{\delta}_{Ei}'\boldsymbol{\lambda}_E \tag{9.5}$$

其中,$\mathbf{d}_i$ 为7维向量,其第 $i$ 个元素等于1,所有其他元素等于0。$\boldsymbol{\delta}_{Ei}$为题器指标5,6及7的3维向量,等于 $\mathbf{d}_i$ 的最后3个元素,如 $\boldsymbol{\delta}_{E6}=(0,1,0)'$。如果不考虑极端情形因子的话,模型在增加3个自由度的情况下偏差增加207.7,因而单维因子模型显得不太适当。

既然对于每个受访者及每个题器均存在重复测量,那么,还可在受访者层次上引入题器的独特因子,即

$$\nu_{itjk} = \mathbf{d}_i'\boldsymbol{\beta} + \eta_{Gjk}^{(3)}\mathrm{d}_i'\boldsymbol{\lambda}_G + \eta_{Ejk}^{(3)}\boldsymbol{\delta}_{Ei}'\boldsymbol{\lambda}_E + \sum_{m=1}^{7}\eta_{Umijk}^{(3)}d_{im} \tag{9.6}$$

在这个模型中,潜变量相互独立。式(9.6)中的最后一项独特因子 $\eta_{Umjk}^{(3)}$ 可被解释为在一些题器的态度上,受访者之间存在的异质性,这些异质性会带来未被公共因子所解释的额外的对不同时期应答的依赖。对该模型对数似然的计算需要在第3层次上对9个维度进行积分。为了降低维度,题器 $i$ 可以被当成第2层次的单元,这样时间就成为第1层次,从而该模型可参数优化为

$$\nu_{itjk} = \mathbf{d}_i'\boldsymbol{\beta} + \eta_{Uijk}^{(2)}\mathbf{d}_i'\boldsymbol{\lambda}_U + \eta_{Gjk}^{(3)}\mathbf{d}_i'\boldsymbol{\lambda}_G + \eta_{Ejk}^{(3)}\boldsymbol{\delta}_{Ei}'\boldsymbol{\lambda}_E \tag{9.7}$$

在此,式(9.6)的最后一项题器 $i$ 的 $\eta_{Uijk}^{(3)}$,被替换为 $\eta_{Uijk}^{(2)}\boldsymbol{\lambda}_{Ui}$。而 $\eta_{Uijk}^{(3)}$ 被当做每个题器 $i=1,\cdots,7$ 的潜变量。$\eta_{Uijk}^{(2)}$ 则是对于不同题器 $i$ 的具有不同实现的单个潜变量。$\boldsymbol{\lambda}_U$ 的目的是为了允许独特因子方差在不同题器中不同。由于 $\eta_{Uijk}^{(3)}$ 与 $\eta_{Uijk}^{(2)}\boldsymbol{\lambda}_{Ui}$ 在题器间变化,与题器不相关,且在每个题器上的方差不同,因此,这两个模型是等价的。式(9.7)的优点在于将第3层次上的9维积分转化为第2层次上的一维积分与第3层次上的二维积分。通常可通过参数优化来将高维积分降维转化为等价模型(见5.3.2节)。增加受访者层次上的独特因子到此双因子模型中,增加了7个参数,偏差降低12.6,仅有微小变化。

对于该二维方差成分因子模型(见4.3节),除了增加受访者层次的潜变量以外,还可再引入区域层次的潜变量,使得公共因子可在不同调查区域变化。允许一般态度因子在不同区域变化,在增加了一个参数的情况下将模型偏差降低8.2,而允许极端情形因子在不同区域变化,仅将模型偏差降低3.2。最终保留的模型为式(9.5)的响应模型上加上结构模型,即

$$\eta_{Gjk}^{(3)} = \eta_{Gk}^{(4)} + \zeta_{Gij}^{(3)}$$

这个最终模型的路径图如图9.10所示。请注意,指向响应变量的路径并非代表在该响应上的线性效应,从 $\eta_G^{(3)}$ 与 $\eta_E^{(3)}$ 发出的路径表示在对数比率上的线性效应,短箭头表示在给定该模型所暗含的概率条件下,响应变量的随机变差。

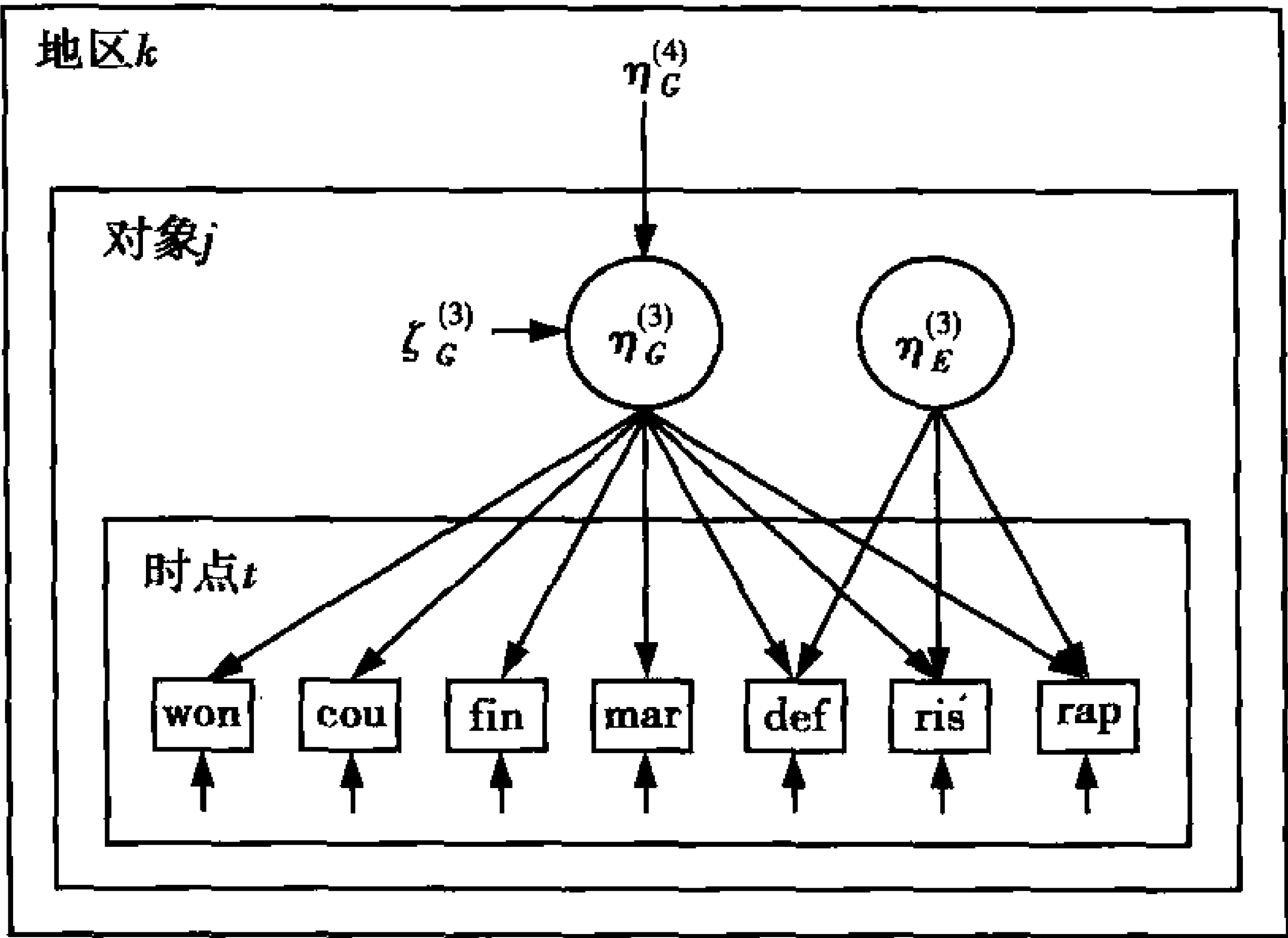



**图 9.10　多层次方差成分因子模型路径图**

如果在区域层次上引入独特因子,第 4 层次的积分维度将从 1 增加到 8。在这个例子中,无法通过参数优化进行降维。因此,我们分别引入独特因子,但是在偏差上的改变很小。对于最终保留的多层次方差成分对数概率因子模型的参数估计结果如表 9.14 所示。计算结果通过在每个维度上采用 10 点自适应积分法得到,在每个维度上采用 5 点或 8 点的计算结果与此也相当接近。正如事先所预期的,由于认可这些题器的发生率很高,极端情形因子的题器截距远大于其他题器。在受访者层次上,包含一般态度因子及特殊情形因子。只有一般态度因子看上去在不同调查区域间有所变化,但其标准差非常小。

**表 9.14　多层次方差成分对数概率因子模型**

| | | |
|---|---|---|
| 固定部分 | | |
| 截距 | | |
| $\beta_1$ [Woman] | -0.83(0.14) | |
| $\beta_2$ [Couple] | -0.17(0.15) | |
| $\beta_3$ [Marriage] | -0.28(0.16) | |
| $\beta_4$ [Financial] | -0.01(0.14) | |
| $\beta_5$ [Defect] | 3.79(0.27) | |
| $\beta_6$ [Risk] | 5.90(0.56) | |
| $\beta_7$ [Rape] | 4.82(0.39) | |
| 随机部分: 受访者层次 | | |
| 因子负荷 | 一般 | 极端 |
| $\lambda_{G1}$ & $\lambda_{E1}$ [Woman] | 1 | 0 |
| $\lambda_{G2}$ & $\lambda_{E2}$ [Couple] | 1.13(0.08) | 0 |
| $\lambda_{G3}$ & $\lambda_{E3}$ [Marriage] | 1.21(0.09) | 0 |
| $\lambda_{G4}$ & $\lambda_{E4}$ [Financial] | 1.01(0.08) | 0 |
| $\lambda_{G5}$ & $\lambda_{E5}$ [Defect] | 0.78(0.09) | 1 |
| $\lambda_{G6}$ & $\lambda_{E6}$ [Risk] | 0.73(0.13) | 1.53(0.26) |
| $\lambda_{G7}$ & $\lambda_{E7}$ [Rape] | 0.72(0.11) | 1.23(0.21) |
| 因子方差 $\psi_G^{(2)}$ & $\psi_E^{(2)}$ | 5.22(0.67) | 3.30(0.80) |
| 随机部分:区域层次 | | |
| 因子方差 $\psi_G^{(3)}$ & $\psi_E^{(3)}$ | 0.36(0.17) | 0 |
| 对数似然值 | -5 160.9 | |

资料来源:Rabe-Hesketh et al. (2004a)

## 9.9 总结与进一步阅读

在本章,我们首先分析了呼吸道感染的纵贯性数据。我们使用了随机截距对数概率回归模型,并与 GEE 模型的结果相比较。关于丛聚二元数据(clustered binary data)分析的评述请参见 Neuhaus (1992) 以及 Pendergast et al. (1996)。关于条件或边缘方法之间的优缺点分析请见 Lindsey and Lambert (1998)、Lindsey (1999)及 Crouchley and Davies (1999)。斯科隆多等(Skrondal et al., 2000)用双层次随机截距模型来研究艾滋病诊断后的避孕套使用变化情况,拉贝-赫斯基思等(Rabe-Hesketh et al., 2001c)在神经心理学中用三层次随机截距模型研究重复测量的精神分裂症患者,他们的健康亲属及其他控制问题。迪歌等(Diggle et al., 2002)、斯科隆多和拉贝-赫斯基思(Skrondal and Rabe-Hesketh, 2003c)、拉贝-赫斯基思和艾弗瑞特(Rabe-Hesketh and Everitt, 2003)以及其他学者用一个双层次随机系数模型来分析关于精神紊乱的纵贯性数据。

9.3 节是关于心肌梗塞的潜类模型分析。这个模型在医学上的其他应用请见福尔曼和科洛曼 (Formann and Kohlmann, 1996)。另外,请参见 13.5 节的关于排序的潜类模型,13.6 节的首选项分析,12.4.4 节的生存分析,以及 14.3 节与 14.4 节的多过程(multiple processes)分析。克劳格(Clogg, 1995)对于潜类模型做了一个很好的综述。

另外,我们还对能力测试进行了包含协变量的单参数、双参数及三参数题器响应模型分析。在 9.8 节讨论了多层次题器响应模型,在下一章将讨论定序响应的多维双参数模型。

关于尼古丁咀嚼胶对戒烟效应的元分析,我们采用了个体层次上的随机效应模型,并对贝叶斯法和似然法估计进行了比较。尽管元分析法在医学上已非常普遍,它在其他学科也取得越来越广的应用,如经济学(如 Granger, 2002)和社会学(如 DiPrete, 2002)。关于元分析的书有 Hedges and Olkin (1985),另外,Whitehead (2002)、Fleiss (1993) 与 Normand (1999)对元分析也提供了很有帮助的综述。

对于妻子的就业转换数据,采用了不同类型的随机效应马尔可夫模型来进行分析,以探究真实状态依赖与虚假状态依赖。在本章中,响应变量被处理为二项变量,在第 12 章中,会将其处理为离散时间存续(discrete time duration)变量。

对于白靴兔种群规模的估计,采用了考虑到未观测到的异质性的捕获-再捕获模型。关于这些方法的综述请见赵莲菊等(Chao et al., 2001)。

最后,关于堕胎态度我们介绍了多层次题器响应模型。福克斯 (Fox, 2001)在一项教学环境研究中探讨了用多层次结构方程模型来分析二项变量。安萨里和杰迪迪(Ansari and Jedidi, 2000) 以及福克斯和格拉斯 (Fox and Glas, 2001)讨论了多层次题器响应模型的贝叶斯估计。

本章中讨论的所有模型都采用了对数概率链接函数,但也可以使用其他形式,如正态概率或余重对数链接函数。

# 10 定序响应

## 10.1 介 绍

本章的第一个主题是关于重复测量的定序响应变量的群组随机试验效应增长曲线模型。首先，我们讨论重复测量的单个定序观测响应变量的多层次增长模型。随后将这些模型扩展到重复测量的定序观测潜变量增长模型。本章的另一个主题是定序测量量表的"心理测量校验(psychometric validation)"。我们将展示如何来研究模型的一些属性，如因子维度、题器信度和题器偏误等。

## 10.2 性教育群组随机试验：潜增长曲线模型

### 10.2.1 介 绍

群组随机试验是指将群组随机分配到试验组或控制组，而非直接对个体进行随机分配。它的一个典型应用为非治疗性试验的评估，如不同方式的性教育对使用避孕用品的效应。

群组随机试验具有许多优点：首先，有些试验是自然应用在群组层次上的。很明显的例子如性教育，性教育通常在学校班级中进行，而将所有学生进行随机分配的做法在实际中是不可行的。其次，群组随机试验减少了试验污染(experimental contamination)。例如，在性教育的例子中，如果接受试验的学生将他的知识与没有接受试验的学生分享，就会带来"污染"。通过在学校层次上进行随机分配，假设不同学校之间的学生不可能相互沟通，可以将污染最小化。

然而群组随机试验也有一些缺点：首先，通常来说，在群组中的个体相比不同群组中的个体更加类似。这意味着在统计模型中不能把个体当做独立的来处理；组内个体的依赖必须加以考虑。其次，与第一点相关，和经典随机试验相比，群组随机试验的效率更低。

在本节中，我们将分析挪威15岁及16岁在校学生关于性教育的群组随机试验数据(Træen, 2003)[1]。性教育通过学校发放的手册及课程教育，包括让学生表演戏剧，进行辩论赛以及一些实践活动，如如何去获得避孕套。这些试验的目的是使青少年在性背景中成为有能力的行动者，敢于面对诸如避孕这样的问题及承担压力。

---

1 感谢本特·特拉恩(Bente Træen)提供数据。

研究关注的并不在于是否使用避孕套,而是一小部分性行为较活跃的青少年。因而,所研究的是关于"避孕的自我效能(contraceptive self-efficacy)"这个假设构念,而非实际的行为。这个构念在以前的研究中被认为能够很好地预测避孕套的使用(如 Kvalem and Træen,2000)。在本节,我们关注其中一个组成构念——"避孕的情景沟通",用 3 个问卷问题来测量:

"当准备发生性关系时,如果我的伙伴和我都没有提到关于避孕的事,那么……"

- [Tell] 我会告诉他/她我没有带避孕用品。
- [Ask] 我会问他/她是否有避孕用品。
- [Get] 我会拿出一个避孕套(如果我有的话)。

这些问题都为 5 级定序回答:

(1)很不符合。

(2)不符合。

(3)一般。

(4)比较符合。

(5)完全符合。

各个学校通过随机的方式被指定为是否采用试验。在进行随机分配之前,先进行一次问卷调查,并在随机分配之后的第 6 个月和第 18 个月再做两次问卷调查。因此,数据就有 3 个层次的结构,波次 $t$ 嵌套于学生 $j$,再嵌套于学校 $k$。总共有 46 所学校,1 184 名学生参加了该研究。其中,只有 570 名学生回答了全部 3 次问卷调查,另有 400 名回答了其中几次问卷调查,还有 114 名从未回答问卷调查。我们在此使用的两个预测变量如下:

- [Treat]虚拟变量,所在的学校是否接受了试验 $x_{1jk}$(是 =1, 否 =0)。
- [Time] 自从随机分配后计算的时间,按 6 个月为周期 $x_{2tjk}$(0,1,3)。

### 10.2.2 增长曲线模型

我们首先针对其中一个题器[Get]来估计一个多层次成比例比率模型(multilevel proportional odds model)。我们允许潜响应的均值依赖于[Time](线性趋势)、[Treat] 以及 [Time] × [Treat],另外,还包括学生 $\zeta_{jk}^{(2)}$ 与学校 $\zeta_k^{(3)}$ 的随机截距。潜响应变量 $y_{tjk}^*$ 在波次 $t$ 对于学生 $j$ 及学校 $k$ 的模型则为

$$
\begin{aligned}
y_{tjk}^* &= \beta_1 x_{1jk} + \beta_2 x_{2tjk} + \beta_3 x_{1jk} x_{2tjk} + \zeta_{jk}^{(2)} + \zeta_k^{(3)} + \epsilon_{tjk} \\
&= \mathbf{x}'_{tjk}\boldsymbol{\beta} + \zeta_{jk}^{(2)} + \zeta_k^{(3)} + \epsilon_{tjk}
\end{aligned}
$$

其中,$\epsilon_{tjk}$符合对数概率分布。所观测到的响应由阈值模型生成:

$$
y_{tjk} = \begin{cases}
1 & \text{当} \quad y_{tjk}^* \leqslant \kappa_1 \\
2 & \text{当} \quad \kappa_1 < y_{tjk}^* \leqslant \kappa_2 \\
3 & \text{当} \quad \kappa_2 < y_{tjk}^* \leqslant \kappa_3 \\
4 & \text{当} \quad \kappa_3 < y_{tjk}^* \leqslant \kappa_4 \\
5 & \text{当} \quad \kappa_4 < y_{tjk}^*
\end{cases}
$$

请注意,在潜响应模型中常数被忽略,使得全部 4 个阈值 $\kappa_s, s = 1,\cdots,4$,都是相等的。

估计结果如表 10.1 所示。奇怪的是,学校的随机截距方差估计值几乎为零。然而,有足够的证据表明学生中间存在未观测到的异质性,其方差估计值为 2.03,对应于潜响应变量的组间残相关系数 0.37。[Time] × [Treat]存在微弱的试验效应,其估计值为

0.17,优比为1.19。因此,每隔6个月,在试验组中定序响应高对低的比率增长的百分比比控制组高19%(高对低的比率指任意高低组合,如回答5对比回答1~4,或者回答4,5对比回答1~3,或者回答3~5对比回答1,2,或者回答2~5对比回答1)。

表10.1 多层次成比例比率模型估计值

| 参数 | 单层次模型 | | 双层次模型 | | 三层次模型 | |
|---|---|---|---|---|---|---|
| | Est | (SE) | Est | (SE) | Est | (SE) |
| 回归系数 | | | | | | |
| $\beta_1$[Time] | -0.12 | (0.06) | -0.13 | (0.06) | -0.13 | (0.06) |
| $\beta_2$[Treat] | -0.05 | (0.14) | -0.02 | (0.19) | -0.02 | (0.19) |
| $\beta_3$[Time]×[Treat] | 0.17 | (0.08) | 0.17 | (0.09) | 0.17 | (0.09) |
| 方差 | | | | | | |
| 学生层次 $\psi^{(2)}$ | — | | 2.03 | (0.31) | 2.03 | (0.31) |
| 学校层次 $\psi^{(3)}$ | — | — | | | 0.00 | (0.00) |
| 阈值 | | | | | | |
| $\kappa_1$ | -3.54 | (0.17) | -4.41 | (0.23) | -4.41 | (0.23) |
| $\kappa_2$ | -2.43 | (0.13) | -3.15 | (0.19) | -3.15 | (0.19) |
| $\kappa_3$ | -1.18 | (0.12) | -1.58 | (0.16) | -1.58 | (0.16) |
| $\kappa_4$ | 0.16 | (0.12) | 0.25 | (0.15) | 0.25 | (0.15) |
| 对数似然值 | -2 531.21 | | -2 470.88 | | -2 470.88 | |

为了直观显示试验效应,图10.1按时间阶段及组别显示响应值为 $s$ 或更高($s=2,3,4,5$)的总体平均概率 $\Pr(y_{tjk} \geqslant s \mid \mathbf{x}_{tjk};\hat{\boldsymbol{\theta}})$。这些概率通过在给定随机效应条件下,将条件响应概率对随机效应分布求积分而得。其相应的观测到的比例也给予显示。值得一提的是,在对数概率尺度上的线性关系并不总是在概率尺度上也呈现出类似这样的线性关系。

### 10.2.3 潜增长曲线模型

既然没有充分的证据说明学校间存在变差,那么就让我们忽略下标 $k$。我们首先分析"避孕的自我效能"的测量模型 $\eta_{tj}$,通过3个定序题器 $y_{itj}$ 进行测量,$i=1$ [Tell],$i=2$ [Ask]以及 $i=3$ [Get]。首先考虑单因子模型,用3种不同的阈值 $\kappa_{is}$ 设定以及截距 $\delta_i$:

- 对每个题器 $i$ 都有不同的阈值设定,无截距(12个参数)。

$$
y_{itj} = \begin{cases} 1 & \text{当} \quad y_{itj}^* \leqslant \kappa_{i1} \\ 2 & \text{当} \quad \kappa_{i1} < y_{itj}^* \leqslant \kappa_{i2} \\ 3 & \text{当} \quad \kappa_{i2} < y_{itj}^* \leqslant \kappa_{i3} \\ 4 & \text{当} \quad \kappa_{i3} < y_{itj}^* \leqslant \kappa_{i4} \\ 5 & \text{当} \quad \kappa_{i4} < y_{itj}^* \end{cases}
$$

- 3个题器具有相同的阈值设定,无截距(4个参数),即

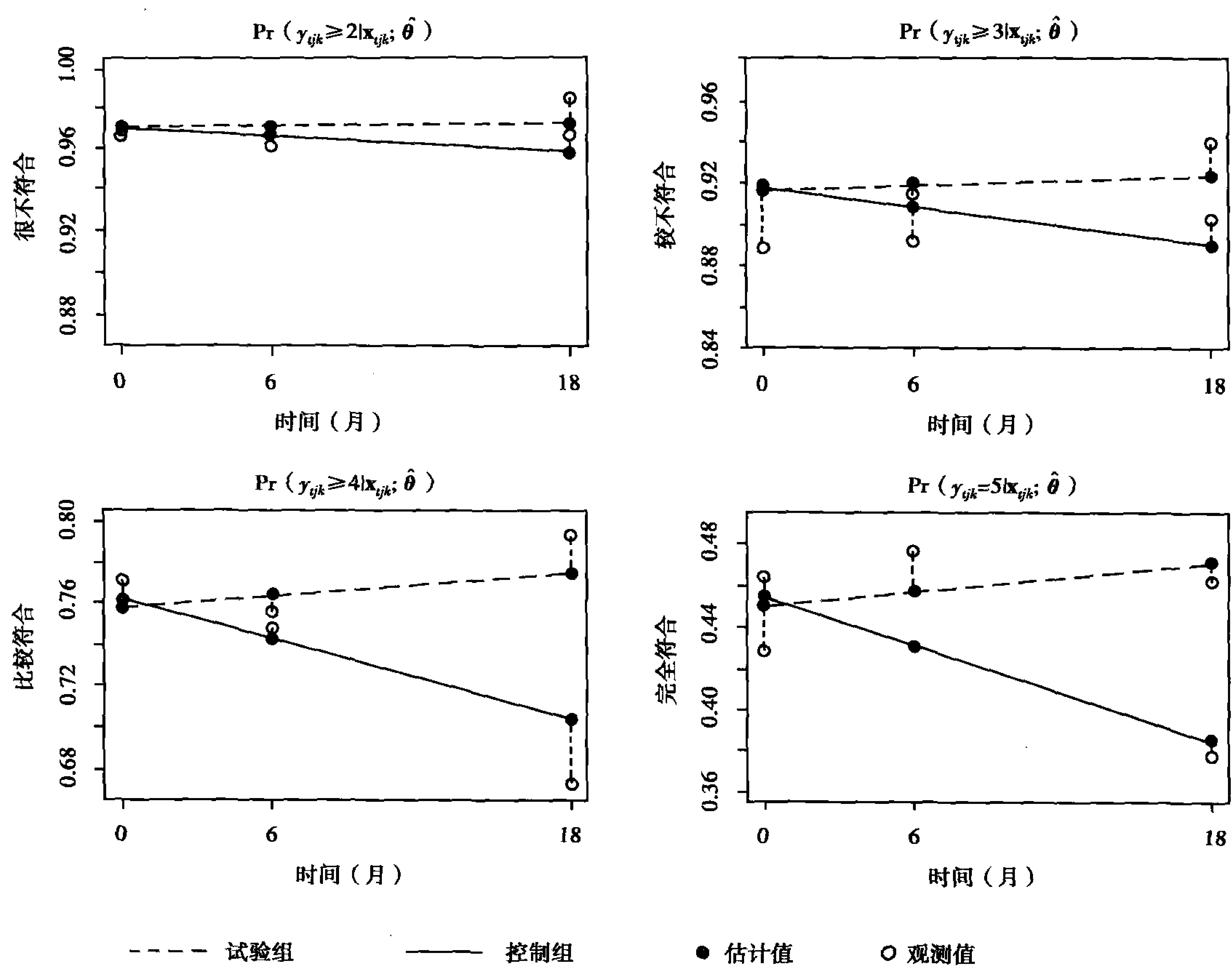



图 10.1　估计的及观测到的边缘响应概率

$$y_{itj}=\begin{cases}1 & \text{当} \quad y_{itj}^{*}\leqslant\kappa_1\\ 2 & \text{当} \quad \kappa_1<y_{itj}^{*}\leqslant\kappa_2\\ 3 & \text{当} \quad \kappa_2<y_{itj}^{*}\leqslant\kappa_3\\ 4 & \text{当} \quad \kappa_3<y_{itj}^{*}\leqslant\kappa_4\\ 5 & \text{当} \quad \kappa_4<y_{itj}^{*}\end{cases}$$

- 3 个题器具有相同的阈值设定，题器 2 截距为 $\delta_2$，题器 3 截距为 $\delta_3$(6 个参数)。

3 个模型的对数似然值分别为 −6 946, −6 990 与 −6 950，故保留最后一个模型。对于在波次 $t$ 中，学生 $j$ 在第 $i$ 个题器上的潜响应 $y_{itj}^{*}$ 因而可用模型表达为

$$y_{itj}^{*}=\delta_i+\lambda_i\eta_{tj}^{(2)}+\epsilon_{itj},\lambda_1=1,\delta_1=0$$

其中，常数阈值 $\kappa_s,s=1,2,3,4$。

我们接下来将测量模型与结构模型进行合并，则

$$\eta_{tj}^{(2)}=\gamma_1x_{1j}+\gamma_2x_{2tj}+\gamma_3x_{1j}x_{2tj}+\eta_j^{(3)}+\zeta_{tj}^{(2)},\qquad \eta_j^{(3)}=\zeta_j^{(3)}$$

其中，$\eta_j^{(3)}$ 为学生层次上的随机截距，$\zeta_{tj}^{(2)}$ 为每隔波次特定的截距。此潜增长曲线模型的路径图如图 10.2 所示，其中 3 个潜变量 $\eta_{1j}^{(2)}$、$\eta_{2j}^{(2)}$ 和 $\eta_{3j}^{(2)}$ 代表 $\eta_{tj}^{(2)}$，$t=1, 2, 3$，对于 $\zeta_{tj}^{(2)}$ 和 $\mathbf{x}_{tj}$ 也用同样的表示法。或者，我们可将 $\eta_{tj}^{(2)}$ 和 $\mathbf{x}_{tj}$ 放到“第 2 层次”的框中，如图10.3 所示。

此潜增长曲线模型的参数估计如表 10.2 所示。试验效应的估计值与前面的结果相

似。奇怪的是，在控制组中“避孕的自我效能”降低，但是在试验组中与事先估计一致，效能增加。在学生之间及学生的不同波次内都存在较大的方差。因为 $\delta_2$ 与 $\delta_3$ 的估计值为正数，故[Ask] 和[Get]看上去要比[Tell]更“容易”。尽管因子负荷非常接近于1，但由于标准误很低，因而不能将它们约束为1。

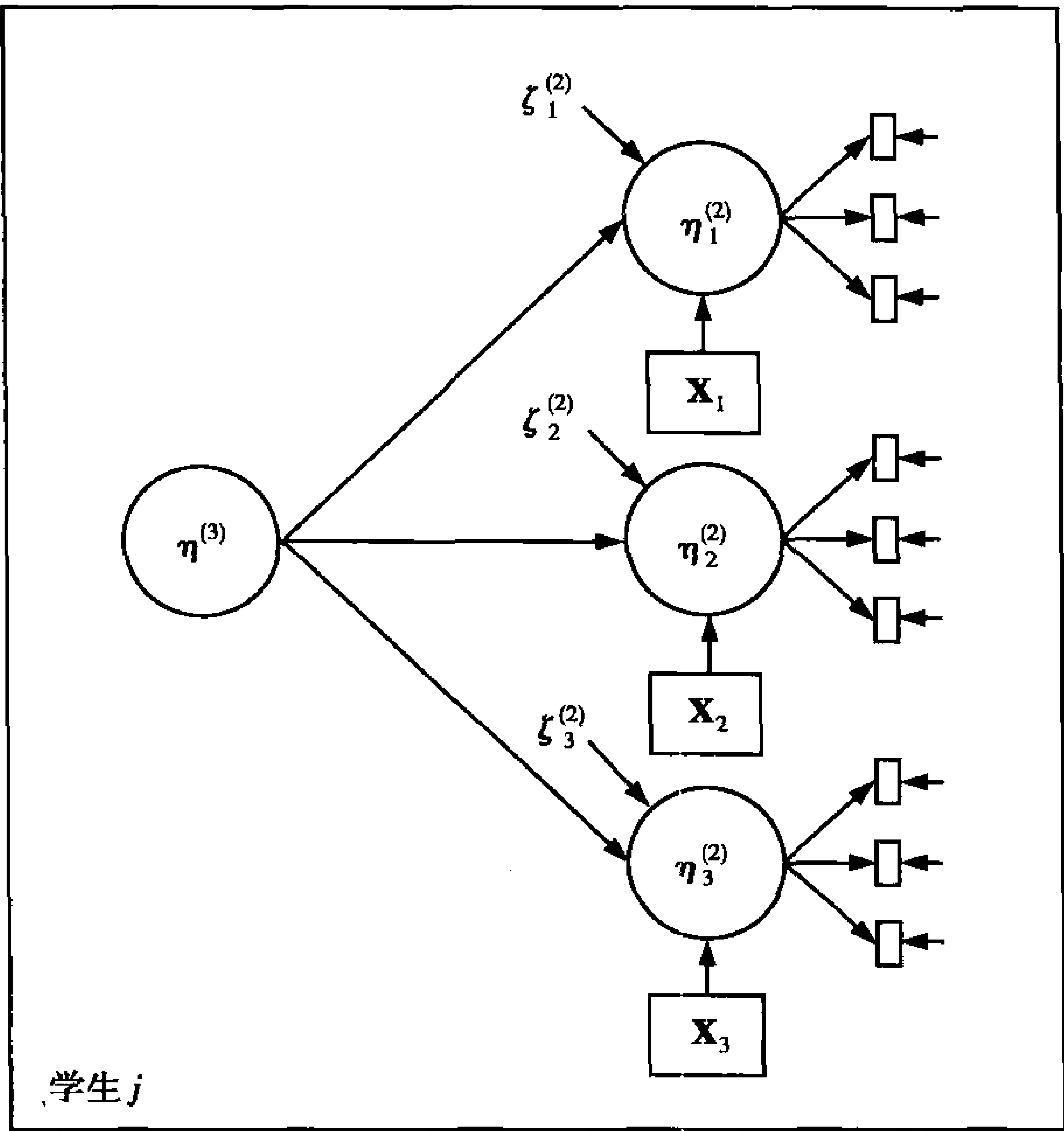



**图 10.2 潜增长曲线模型路径图**

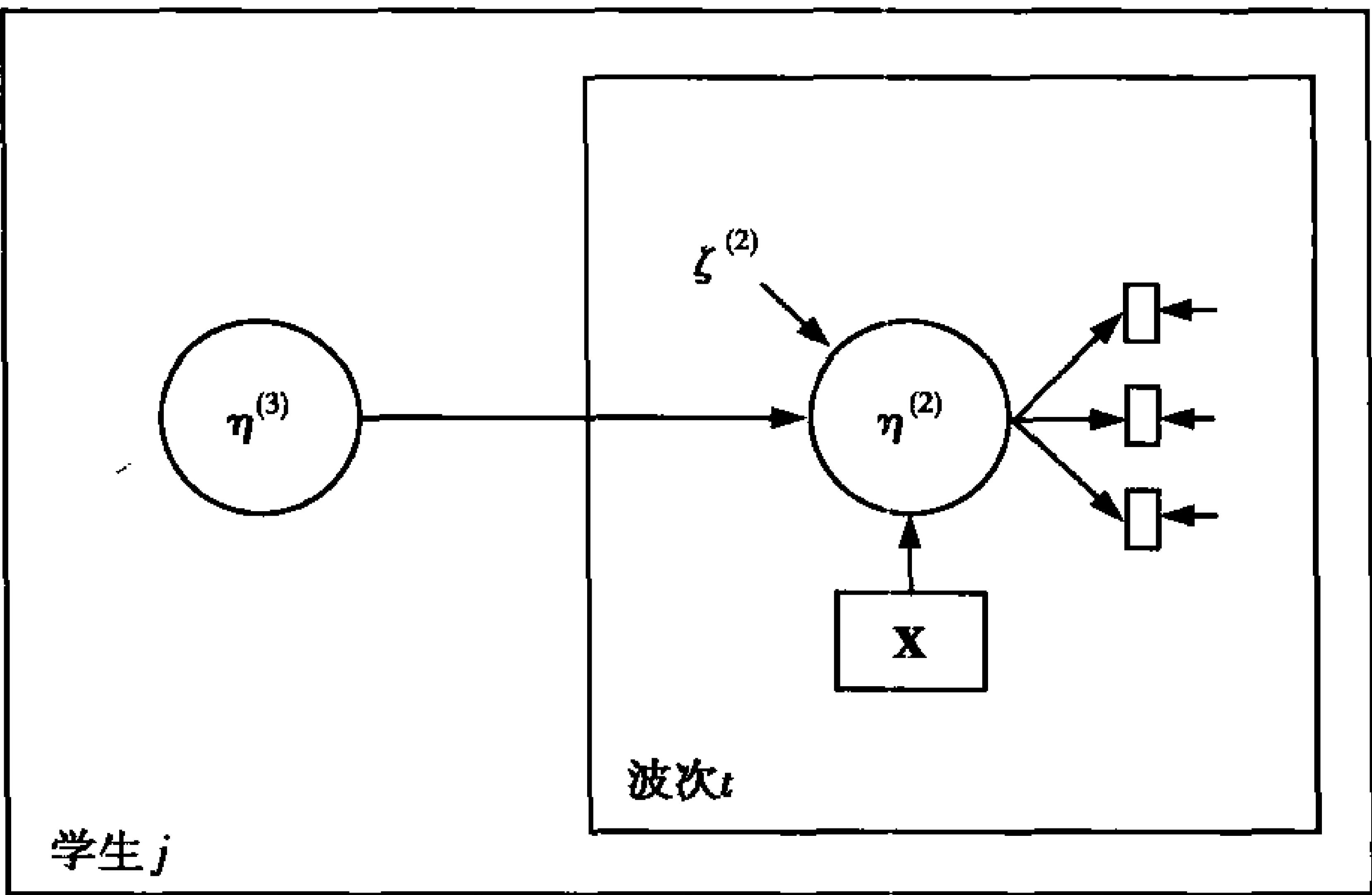



**图 10.3 另一种潜增长曲线模型路径图**

**表 10.2 潜增长曲线模型参数估计值**

| 参　数 | Est | (SE) |
|---|---|---|
| 结构模型 | | |
| 回归系数 | | |
| $\gamma_1$[Time] | -0.06 | (0.09) |
| $\gamma_2$[Treat] | -0.28 | (0.24) |
| $\gamma_3$[Time]×[Treat] | 0.20 | (-0.11) |

续表

| 参　数 | Est | (SE) |
|---|---|---|
| 方差 | | |
| 波次层次 | | |
| $\psi^{(2)}$ | 4.57 | (0.41) |
| 学生层次 | | |
| $\psi^{(3)}$ | 3.72 | (0.43) |
| 测量模型 | | |
| 截距 | | |
| $\delta_1$[Tell] | 0 | |
| $\delta_2$[Ask] | 0.42 | (0.08) |
| $\delta_3$[Get] | 0.52 | (0.08) |
| 因子负荷 | | |
| $\lambda_1$[Tell] | 1 | |
| $\lambda_2$[Ask] | 1.09 | (0.04) |
| $\lambda_3$[Get] | 0.91 | (0.04) |
| 阈值 | | |
| $\kappa_1$ | -5.46 | (0.24) |
| $\kappa_2$ | -3.74 | (0.23) |
| $\kappa_3$ | -1.82 | (0.22) |
| $\kappa_4$ | 0.87 | (0.21) |
| 对数似然值 | -6 860.96 | |

## 10.3 政治效能：因子维度及题器偏误

### 10.3.1 介　绍

政治效能(Political efficacy)是政治态度、行为及参与理论研究中的重要概念。这个假设的概念首先由密歇根大学社会调查研究中心(Survey Research Center of the University of Michigan)于1952年提出并予以操作化(Campbell et al., 1954)。坎贝尔等(Campbell et al.:187)将政治效能定义为：

……认为个体的政治行动能够对政治过程产生影响的信念……

认为政治或社会变迁是可能的，并且个体公民能够参与到该变迁过程当中。

最初，该概念被操作化为4个二项题器。1968年，又附加了两个题器到量表当中，

- [Nosay] 我以及其他和我一样的人对政府的行为是没有发言权的。
- [Voting] 投票是我以及其他和我一样的人对政府如何运作发表看法的唯一方式。
- [Complex]政治或政府太复杂，我有时候无法完全明白他们在干什么。

• [Nocare] 我不认为政府官员会在乎我们的看法。

• [Touch] 总的来讲，被我们选到华盛顿国会去的那些人很快就不再与选民们联系了。

• [Interest] 政党只关心选民的选票，而非他们的意见。

另外，二项响应类别被替换成四级定序类别：

(1)非常不同意。

(2)不同意。

(3)同意。

(4)非常同意。

这个效能量表已被用于多项经验研究中。其中一个例子为 8 国调查的"政治行动数据"[1] (Wieken-Mayser et al. , 1979;另见 Barnes et al. , 1979)。

根据斯科隆多 (Skrondal, 1996)，我们将分析 1974 年美国子样本的 1 719 个被访者数据。出于演示的目的，我们将回答"不知道"的当做缺失值处理(另见 Rubin et al. , 1995)。效能题器的单变量频次分布如表 10.3 所示，带有缺失值的题器的频次分布如表 10.4 所示。在随机性缺失的假设下，共有 1 710 个回答了至少一个效能题器的被访者进入分析数据。

**表 10.3 政治效能题器单变量频次分布**

| | 4 | 3 | 2 | 1 | 缺失 |
|---|---|---|---|---|---|
| [Nosay] | 175 | 518 | 857 | 130 | 39 |
| [Voting] | 283 | 710 | 609 | 80 | 37 |
| [Complex] | 343 | 969 | 323 | 63 | 21 |
| [Nocare] | 250 | 701 | 674 | 57 | 37 |
| [Touch] | 273 | 881 | 462 | 26 | 77 |
| [Interest] | 264 | 762 | 581 | 31 | 81 |

**表 10.4 带缺失值的题器频次统计**

| 缺失题器个数 | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| 频次 | 1 554 | 106 | 26 | 18 | 4 | 2 | 9 |

## 10.3.2 因子维度及信度

让我们首先分析政治效能题器的因子维度。为此，我们将使用如下形式的定序正态概率因子模型：

$$\mathbf{y}_j^* = \mathbf{\Lambda}\boldsymbol{\eta}_j + \boldsymbol{\epsilon}_j$$

由于忽略了常数项，因此对于每个题器 $i$，可以辨识出全部 4 个阈值 $\kappa_{1i},\cdots,\kappa_{4i}$。对于单维定序正态概率因子模型的情形，我们可以得到在题器响应理论中所谓的评分响应

---

1 本节所用数据由巴恩斯(S. H. Barnes)、英格尔哈特(R. Inglehart)、詹宁斯(M. K. Jennings)以及法拉(B. Farah)处理，来自挪威社会科学数据中心(Norwegian Social Science Data Services (NSD))。该数据的收集者及 NSD 均不对本书的分析结果负责。

(graded response) 模型(Samejima, 1969)。

单维因子模型对于**单维性**(unidimensionality)概念提供了形式化表达,与应用科学家的想法不谋而合(McDonald, 1981)。如果要表示多维性(multidimensionality),则直观性要差些。例如,可以表示大多数符合**二维性**(bidimensionality)概念的因子模型。在本节,我们将详细解说4种类型的**因子二维性**(factor bidimensionality),按其严格程度从高到低排列:

(1)**严格因子二维性**(Strict factor bidimensionality)被形式化为一个**独立群组**因子模型,模型的题器仅测量它们所要测量的维度。也就是说,$\Lambda$ 被设定为块角结构。因子维度被先验设定为**正交**;因子 $\boldsymbol{\psi}$ 的协方差矩阵为对角矩阵。如果保留该模型,那么题器的维度效度(dimensional Validity)最大化。

(2)**强因子二维性**(Strong factor bidimensionality)也同样被形式化为一个独立群组因子模型,但是因子之间允许相关(见图10.4上图)。所以,强二维性概念比严格二维性要弱一些。

(3)当有一个或多个题器(但不是全部)测量两个维度时,就称为**中度因子二维性**(Intermediate factor bidimensionality)。这种组合的题器通常存在问题,因为它们测量了不同的现象。

(4)**弱因子二维性**(Weak factor bidimensionality)对应于无约束的或探索性因子模型(Jöreskog, 1969; Lawley 与 Maxwell, 1971)。所有的题器都可反映政治效能的两个维度(在满足模型辨识的前提下)。因此,这是二维性在因子模型里所能实现的最弱的形式化表达。如果该模型获得保留,题器的维度效度最低。关于如何设定探索性因子模型使其等效于验证性因子模型在3.3.3节中已讨论过。对于探索性双因子模型的参数化请见图10.4的下图。

已有大量的文献讨论政治效能题器的测量属性,并且很多学者提出了不同的替代性测量模型。尽管有些学者认为效能为单维性概念,但绝大多数人偏向于将其看做二维性概念。然而在具体到哪个题器测量哪一个维度时,则存在激烈的争论。

在此,我们将讨论局限于"全国竞选研究"(NES)模型(Miller et al., 1980)。米勒等(Miller et al., 1980:253)对于政治效能的两个维度提出了如下的解释:一个维度被解释为"个体对其能够理解政治并有能力参与政治行动(如投票)的自我感知",另一个维度为"个体对于政治制度的信心,而非其对自身能力的感知"。第一个维度被命名为"内部效能"(或个人政治能力),第二个维度被称为"外部效能"(或政治系统支持度)。米勒等认为[Nosay]、[Voting]及[Complex]测量内部效能,而[Nocare]、[Touch]及[Interest]测量外部效能。因此,将NES模型看做是强二维性模型是有道理的(请见图10.4中的上图)。

到此为止,有必要指出在关于如何判定因子模型维度的问题上很多文献经常采用一些肤浅的做法。出于很多理由,务必要小心处理这个问题。首先,绝对拟合的检验问题是造成模型和数据互相矛盾的来源。拟合不佳并非只是因为维度设定不正确,可能还是因子模型设定不正确的结果。对于定类数据来说,这个问题就变得更加复杂,因为阈值的标量设定将影响绝对拟合。其次,正如在8.5.2节所指出的,关于拟合优度存在多个可能互相矛盾的标准。最后,必须处理等价问题。通过对政治效能的文献研究可以说明这个问题。克雷格和与马吉奥多 (Craig and Maggiotto, 1982)设定了一个双因子模型,其中[Nosay]、[Nocare]、[Touch]及[Interest]测量一个维度,[Voting]与[Complex]测量另一个维度。相反,梅森等(Mason et al., 1985)设定了一个三因子模型,将克雷格和马吉

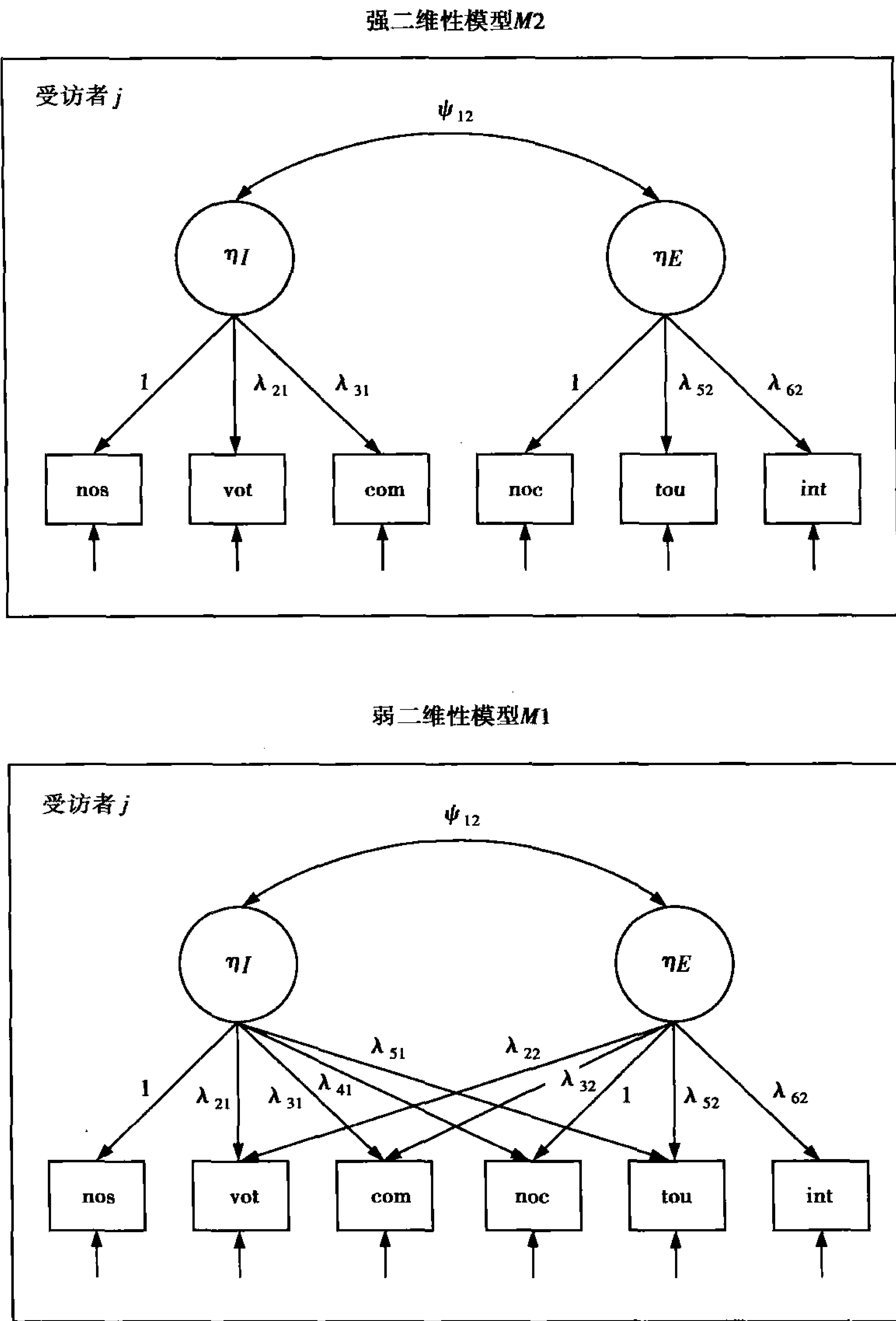



**图 10.4 模型 M2 与 M1 的图形展示**

奥多的第 2 个因子拆分成两个因子,分别用[Voting]与[Complex]测量。这两个模型都可以辨识并且等价,尽管第 2 个模型在辨识上有些勉强。然而,没有经验证据能够帮助确定选择这两个关于政治效能概念的二维和三维模型(见 5.1 节)。

暂且放下上面这个问题,我们现在将这个 NES 模型 *M*2 与弱维度性模型 *M*1 进行比较。这两个模型如图 10.4 所示。我们注意到,*M*2 嵌套于 *M*1,因为 *M*2 中的因子负荷 $\lambda 41$, $\lambda 51$, $\lambda 22$ 以及 $\lambda 32$ 在 *M*2 中被设置为 0。所以,似然比值检验可用于比较这两个模型的拟合。在二元正态潜特质的设定下使用所有的数据,可得到不同模型的最大似然估计值。所估计的因子负荷、因子方差与协方差如表 10.5 所示,阈值如表 10.6 所示。

政治效能的这两个维度存在高度的相关(*M*2 模型为 0.84,*M*1 模型为 0.76),在图 10.5 的经验贝叶斯或因子得分图形上也同样能看出。比较模型 *M*2 与 *M*1,似然比值统计量为 52.79,4 个自由度。我们可知,强二维性对于政治效能题器是明显不合适的,弱二维性模型得到保留。因此,效能题器的维度效度较低。顺便提一下,我们注意到一些模型采用中度二维性(如 Aish and Jöreskog, 1989),但在此不对这些模型继续分析。

**表 10.5　估计的因子负荷及方差(协方差)**

| | *M*2 | | *M*1 | |
|---|---|---|---|---|
| | 内部 | 外部 | 内部 | 外部 |
| 因子负荷 | | | | |
| $\lambda_{1k}$[Nosay] | 1 | 0 | 1 | 0 |
| $\lambda_{2k}$[Voting] | 0.52(0.05) | 0 | 0.69(0.10) | -0.18(0.10) |
| $\lambda_{3k}$[Complex] | 0.77(0.07) | 0 | 0.56(0.08) | 0.15(0.08) |
| $\lambda_{4k}$[Nocare] | 0 | 1 | 0.72(0.12) | 1 |
| $\lambda_{5k}$[Touch] | 0 | 0.74(0.05) | -0.09(0.17) | 1.41(0.26) |
| $\lambda_{6k}$[Interest] | 0 | 0.86(0.06) | 0 | 1.53(0.16) |
| 因子方差(协方差) | | | | |
| $\psi_{kk}$ | 0.81(0.09) | 2.67(0.31) | 0.91(0.13) | 1.02(0.21) |
| $\psi_{12}$ | 1.24(0.10) | | 0.73(0.10) | |
| 对数似然值 | -9 950.43 | | -9 924.04 | |

**表 10.6　模型 *M*2 与 *M*1 所估计的阈值**

| | $\kappa_{i1}$ | $\kappa_{i2}$ | $\kappa_{i3}$ |
|---|---|---|---|
| *M*2 | | | |
| [Nosay] | -1.70(0.06) | -0.28(0.04) | 1.89 (0.07) |
| [Voting] | -1.06(0.04) | 0.25(0.04) | 1.84 (0.06) |
| [Complex] | -1.01(0.05) | 0.91(0.04) | 2.13 (0.07) |
| [Nocare] | -2.00(0.10) | 0.35(0.06) | 3.36 (0.15) |
| [Touch] | -1.53(0.07) | 0.86(0.06) | 3.27 (0.12) |
| [Interest] | -1.69(0.07) | 0.58(0.06) | 3.50 (0.15) |
| *M*1 | | | |
| [Nosay] | -1.74(0.08) | -0.29(0.05) | 1.94 (0.08) |
| [Voting] | -1.08(0.04) | 0.26(0.04) | 1.90 (0.07) |
| [Complex] | -0.99(0.05) | 0.90(0.04) | 2.09 (0.07) |
| [Nocare] | -1.96(0.10) | 0.35(0.06) | 3.30 (0.15) |
| [Touch] | -1.63(0.09) | 0.91(0.07) | 3.48 (0.16) |
| [Interest] | -1.80(0.10) | 0.62(0.06) | 3.72 (0.20) |

接下来请考虑保留模型的题器信度。按照杰拉斯科格(Jöreskog, 1971b)以及沃兹等(Werts et al., 1974),我们定义 $\rho_i$ 为潜响应变量 $y_i^*$ 的"信度":

$$\rho_i = \frac{\mathbf{\Lambda}_i \mathbf{\Psi} \mathbf{\Lambda}_i'}{\mathbf{\Lambda}_i \mathbf{\Psi} \mathbf{\Lambda}_i' + \theta_{ii}} \qquad (10.1)$$

其中,$\mathbf{\Lambda}_i$ 为因子负荷矩阵的第 $i$ 行,$\theta_{ii} = 1$ 为本节所估计的正态概率模型。正如我们在

3.3节所讲,除非题器 $i$ 的方差被假设为0,否则 $\rho_i$ 可以被认为是真实信度的**下限**。模型 $M1$ 的内部效能题器[Nosay]、[Voting]及[Complex]信度下限估计分别为0.48,0.23及0.30;外部效能题器[Nocare]、[Touch]及[Interest]信度下限估计分别为0.72, 0.65及0.71。内部效能题器的信度下限值较低。

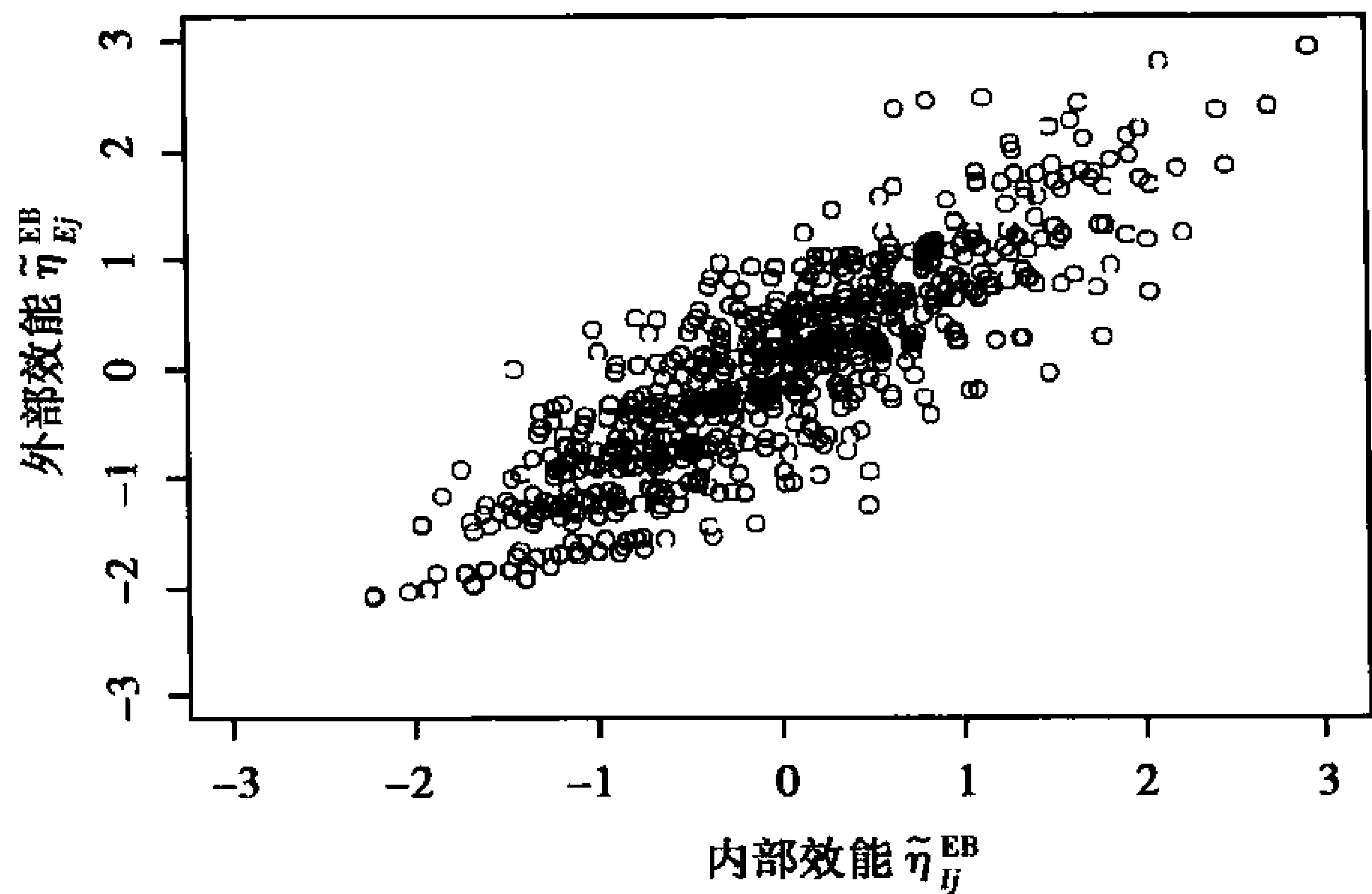



**图10.5 政治效能的经验贝叶斯因子得分($M1$)**

## 10.3.3 题器偏误

我们现在开始分析政治效能题器的**题器偏误**(item bias)或**题器功能差异**(differential item functioning,DIF)。在题器响应理论中,当对题器的响应除了依赖于因子以外,还依赖于外部信息时,则称为"偏误"(另见9.4节关于二项题器响应模型的题器偏误)。题器偏误与测试的"公平性"紧密相关,因而,毫不奇怪,在关于种族的偏误问题上经常导致激烈的公共辩论甚至法律诉讼。我们关于题器偏误的分析基于穆森(Muthén, 1985, 1988b, 1989d)提出的校验方法。根据汉布尔顿和斯瓦米纳坦(Hambleton and Swaminathan, 1985)及汉布尔顿等(Hambleton et al., 1991)的综述,我们认为这个方法比题器响应理论中的标准方法更好、更直接。

我们将模型设定为

$$\mathbf{y}_j^* = \mathbf{X}_j\boldsymbol{\beta} + \boldsymbol{\Lambda}\boldsymbol{\eta}_j + \boldsymbol{\epsilon}_j$$
$$\boldsymbol{\eta}_j = \boldsymbol{\Gamma}\mathbf{w}_j + \boldsymbol{\zeta}_j$$

首先考虑一下不存在题器偏误的模型,$\boldsymbol{\beta}=\mathbf{0}$,即MIMIC模型,效能因子对协变量回归。根据该模型设定,因子的期望值则为异质性的,且响应概率不再为同质性的。因为所有的异质性通过因子传递,所以题器为非偏误的。根据之前对因子维度性的分析结果,我们设定一个非结构化的因子模型(unstructured factor model)。根据艾布拉姆森(Abramson, 1983)及利斯塔乌戈(Listhaug, 1989)的讨论,我们选择两个协变量:

- [Educ] 标准化的受教育年限。
- [Black] 是否为黑人,虚拟变量。

该MIMIC模型用 $M4$ 表示,根据一项初步研究,我们设定其 $\gamma_{22}=0$([Black]对外部效能没有效应)。去除3个在[Black]与[Educ]上存在缺失值的案例,最终得到样本数量1 707个。

现在我们将考虑题器偏误的模型设定为**广义**MIMIC模型(generalized MIMIC),协变量除了通过因子的间接效应以外,对部分题器还具有直接的回归效应。对于模型参数而言,这意味着 $\boldsymbol{\beta}$ 的所有元素都不为0。这个模型用 $M3$ 表示。请注意,$M3$ 并非被当做先

验模型,相反它是根据对数据的探索性分析所做的尝试。通过简单交叉验证(见 8.5.4 节),样本随机地被分到一个大小为 840 的探索性样本及大小为 867 的验证性样本中。从而,我们可以自由地研究探索性样本,然后再将模型在验证性样本中进行检验。探索性研究显示[Educ]与[Black]对[Voting]($\beta_1$ 与 $\beta_2$)存在直接效应,[Educ]对[Complex]($\beta_3$)也存在直接效应。模型 *M*3 和 *M*4 的路径图如图 10.6 所示。

**表 10.7 MIMIC 模型估计值**

| | *M*4 | | *M*3 | |
|---|---|---|---|---|
| | 内部 | 外部 | 内部 | 外部 |
| 结构模型 | | | | |
| 因子回归 | | | | |
| $\gamma_{k1}$[Educ] | 0.34(0.04) | 0.24(0.08) | 0.38(0.07) | 0.28(0.06) |
| $\gamma_{k2}$[Black] | -0.25(0.08) | 0 | -0.37(9.16) | 0 |
| 因子方差(协方差) | | | | |
| $\Psi_{kk}$ | 0.67(0.05) | 0.92(0.23) | 0.97(0.17) | 1.05(0.15) |
| $\Psi_{12}$ | 0.73(0.09) | | 0.69(0.07) | |
| 测量模型 | | | | |
| 因子负荷 | | | | |
| $\lambda_{1k}$[Nosay] | 1 | 0 | 1 | 0 |
| $\lambda_{2k}$[Voting] | 1.67(0.45) | -0.85(0.31) | 0.33(0.10) | 0.04(0.10) |
| $\lambda_{3k}$[Complex] | 1.64(0.41) | -0.60(0.26) | 0.20(0.09) | 0.32(0.10) |
| $\lambda_{4k}$[Nocare] | 0.82(0.31) | 1 | 0.63(0.18) | 1 |
| $\lambda_{5k}$[Touch] | -0.01(0.30) | 1.34(0.44) | -0.11(0.18) | 1.31(0.30) |
| $\lambda_{6k}$[Interest] | 0 | 1.79(0.43) | 0 | 1.59(0.25) |
| 题器回归 | | | | |
| $\beta_1$[Voting] × [Educ] | 0 | | 0.20(0.05) | |
| $\beta_2$[Voting] × [Black] | 0 | | -0.34(0.14) | |
| $\beta_3$[Complex] ×[Educ] | 0 | | 0.28(0.04) | |
| 对数似然值 | -4 978.08 | | -4 973.36 | |

验证性样本的参数估计如表 10.7 所示,其阈值估计如表 10.8 所示。我们注意到[Black]与内部效能呈负相关,而[Educ]与这两种政治效能都为正相关,与预期相符。由于模型 *M*4 嵌套于 *M*3 中,可根据验证性样本进行似然比值检验。根据表10.7,似然比值统计量为 9.43,3 个自由度,说明包含题器偏误的模型 *M*3 应该予以保留,尽管此证据并不是非常强。由表 10.7 可知,内部效能的两个题器存在实质性的题器偏误。[Black]在[Voting]上存在负的直接效应,[Educ]在[Voting]及[Complex]上存在正的直接效应。

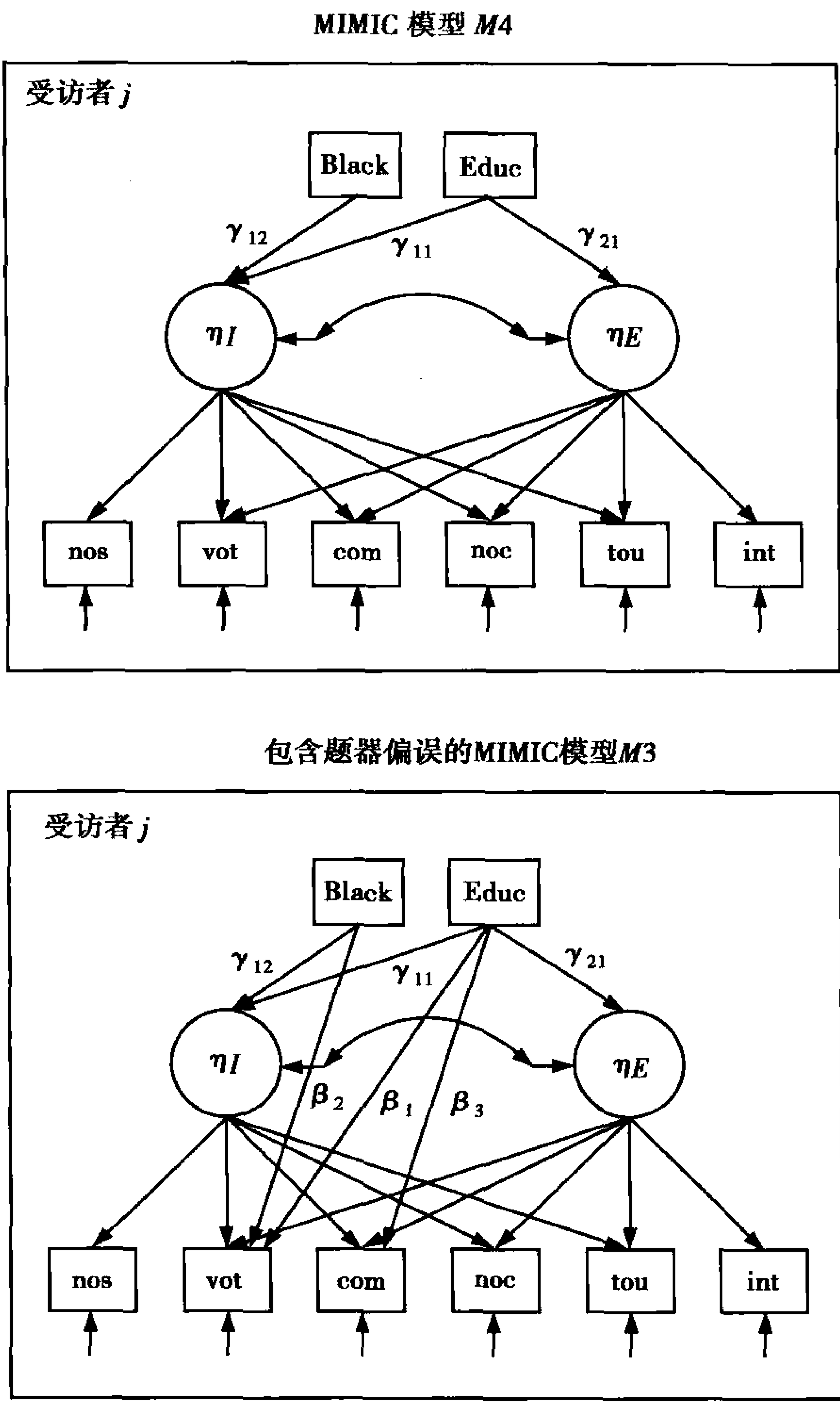



图 10.6　模型 *M*4 与 *M*3 的图形表示

表 10.8　估计的阈值

| | $\kappa_{1i}$ | $\kappa_{2i}$ | $\kappa_{3i}$ |
|---|---|---|---|
| *M*4 | | | |
| [Nosay] | -1.51(0.08) | -0.32(0.06) | 1.74 (0.08) |
| [Voting] | -1.15(0.07) | 0.26(0.05) | 1.92 (0.10) |
| [Complex] | -1.04(0.07) | 0.92(0.06) | 2.09 (0.11) |
| [Nocare] | -1.85(0.12) | 0.30(0.08) | 3.28 (0.17) |
| [Touch] | -1.52(0.10) | 0.85(0.08) | 3.19 (0.16) |
| [Interest] | -1.77(0.13) | 0.66(0.09) | 3.88 (0.27) |
| *M*3 | | | |
| [Nosay] | -1.75(0.17) | -0.38(0.08) | 2.02 (0.19) |
| [Voting] | -1.10(0.06) | 0.24(0.05) | 1.82 (0.09) |
| [Complex] | -0.95(0.06) | 0.90(0.06) | 2.01 (0.09) |
| [Nocare] | -1.92(0.14) | 0.31(0.09) | 3.39 (0.20) |
| [Touch] | -1.57(0.12) | 0.89(0.09) | 3.32 (0.20) |
| [Interest] | -1.80(0.15) | 0.66(0.09) | 3.90 (0.29) |

### 10.3.4 结 论

综上所述,对政治效能题器的心理测量校验尚存在某些不确定的地方。最终保留了弱二维性的假设,信度看上去偏低,并且发现实质性的题器偏误。对于内部效能题器而言,这两个问题混合在一起。所发现的这些问题很可能是由于测量与理论发展之间的概念鸿沟所造成(Mason et al. , 1985)。对于本节所分析的政治效能的 NES 测量量表,我们认为应该放弃。有趣的是,我们也注意到在该领域一个早期的政治家康弗斯 (Converse, 1972:334)也认为:

> 我们从 1952 年以来所使用的政治效能量表包含了许多混杂的东西……

最后,我们看到 NES 于 1988 年使用了新的政治效能量表(如 Niemi et al. , 1991)。

## 10.4 生活满意度:定序有衡正态概率因子模型 (Ordinal scaled probit factor model)

### 10.4.1 介 绍

生活满意度无论在形式上还是在经验层次上都得到非常多的关注。在经验研究中,关于生活满意度最常见的方法是基于人们对满意度感知的自评(如 Campbell et al. , 1976; Andrews and Withey, 1976)。

在本节,我们提供一个关于美国人对生活满意度感知的自评经验研究,最早出现于斯科隆多 (Skrondal, 1996)。该方法论允许我们研究生活满意度的维度以及在具体题器上的性质。在获得一个生活满意度的保留模型之后,我们将模型的属性按拉扎斯菲尔德(Lazarsfeld, 1950)提供的一个方法用图形的方式展示出现。

所使用的数据来自 1989 年的一般社会调查(GSS)[1]。GSS 是对美国年龄在 18 岁及以上居民的横截面社会调查(NORC, 1989)。从 1972 年以来,几乎每年由芝加哥大学的全国民意研究中心进行一次 GSS 调查。GSS 调查的目的为监测社会在态度和行为上的趋势。关于生活满意度的题器问题为:

> 对于下述生活中的一些领域,请告诉我哪个数字反映了您对该领域的满意程度。

受访者对 5 个领域进行评估:

- [City] 所居住的城市。
- [Hobby] 你工作以外的活动——兴趣等。
- [Family] 你的家庭生活。
- [Friend] 你的朋友。
- [Health] 你的健康状况。

受访者对上述题器的评估用下列数字表示:

(1)非常好。

1 本小节所用的数据来自全国民意研究中心(NORC),通过挪威社会科学数据中心发布。全国民意研究中心及挪威社会科学数据中心均不对本分析负责。

(2)很好。
(3)较好。
(4)略好。
(5)一般。
(6)不太好。
(7)很不好。

这种评分法也称为里克特态度量表(Likert,1932)。

从 1988 年起,GSS 采用**分投选票**(split-ballot)的方式进行(Smith, 1988)。具体来说,即部分问题在所有调查中永久性出现,而其他问题包括生活满意度题器则每隔 3 年轮换一次。我们将分析限制在 1989 年的 1 035 个回答了生活满意度问题的受访者数据上。每个题器的频次分布如表 10.9 所示。题器缺失值数量的频次分布如表 10.10 所示。回答了至少一个问题的共有 1 030 个受访者数据,在随机性缺失的假设下进行分析。我们注意到在 1989 年的 GSS 数据中,关于生活满意度方面仅有很少的缺失值。

**表 10.9 生活满意度题器频次分布**

| | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 缺失 |
|---|---|---|---|---|---|---|---|---|
| [City] | 178 | 283 | 217 | 196 | 63 | 69 | 22 | 2 |
| [Hobby] | 243 | 368 | 178 | 109 | 40 | 53 | 34 | 5 |
| [Family] | 433 | 344 | 102 | 67 | 22 | 44 | 15 | 3 |
| [Friend] | 343 | 397 | 156 | 78 | 14 | 32 | 8 | 2 |
| [Health] | 260 | 324 | 174 | 170 | 30 | 48 | 22 | 2 |

**表 10.10 题器缺失值数量频次分布**

| 题器缺失数 | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| 频次 | 1 016 | 14 | 0 | 0 | 0 | 5 |

可设定一个评分响应模型,使得对于每个题器 $i$ 得到下述无约束阈值模型:

$$y_{ij} = \begin{cases} 1 & \text{当} \quad y_{ij}^* \leqslant \kappa_{i1} \\ 2 & \text{当} \quad \kappa_{i1} < y_{ij}^* \leqslant \kappa_{i2} \\ 3 & \text{当} \quad \kappa_{i2} < y_{ij}^* \leqslant \kappa_{i3} \\ 4 & \text{当} \quad \kappa_{i3} < y_{ij}^* \leqslant \kappa_{i4} \\ 5 & \text{当} \quad \kappa_{i4} < y_{ij}^* \leqslant \kappa_{i5} \\ 6 & \text{当} \quad \kappa_{i5} < y_{ij}^* \leqslant \kappa_{i6} \\ 7 & \text{当} \quad \kappa_{i6} < y_{ij}^* \end{cases}$$

共有 30 个阈值参数。

由于响应的类别数较多,这个模型的参数数量特别多。因而,克劳格(Clogg, 1979)以及马斯特斯(Masters, 1985)采用将类别合并的方法。但蒂森和斯滕伯格(Thissen and Steinberg, 1988)对他们的类别合并提出了批评,并提出一个更为合理的合并方案。然而,因为全部题器都使用相同的里克特量表,从而全部题器具有同质性,所以对这个例子来说还有更为合理的方案。在这个例子里,可参照 10.2.2 节,将所有题器的阈值约束为

相等,并对每个题器引入截距,即

$$y_{ij}^{*} = \beta_i + \lambda_i \eta_j + \epsilon_{ij}, \quad \kappa_{is} = \kappa_s, \quad \kappa_1 = 0$$

因为 $\kappa_1$ 被设定为0,故所有题器的截距都是可辨识的。如同5.2.3节所示,将所有题器的阈值设为相等,不仅可以辨识截距,还可在给定有衡定序正态概率模型的条件下得到潜响应变量的相对尺度(参见2.3.4节),即

$$\epsilon_{ij} \sim N(0,\theta_{ii}), \quad \theta_{11} = 1$$

由此,潜响应变量在不同题器中间具有不同的原点及尺度。

## 10.4.2 因子维度

接下来,我们开始分析该GSS调查的5个生活满意度题器。我们的分析与穆拉齐(Muraki, 1990)相同,但与克劳格(Clogg, 1979)、马斯特斯(Masters, 1985)以及蒂森和斯滕伯格(Thissen and Steinberg, 1988)相反,他们将分析限制在其中3个题器上面。另外,克劳格(Clogg, 1988)分析了其中4个题器。

首先我们设定一个单维模型 *M*1,具有如下辨识条件约束:

$$\kappa_1 = 0, \quad \lambda_1 = 1, \quad \theta_{11} = 1$$

所估计的参数及标准误如表10.11所示的第2栏。对估计参数检查结果表明,$\Theta$ 的全部对角线元素的值非常接近,除了 $\theta^{44}$。因而,我们接下来根据 *M*1 对 *M*2 模型进行如下约束:

$$\theta_{22} = \theta_{33} = \theta_{55} = 1 \tag{10.2}$$

*M*2 模型的参数估计及标准误如表10.11所示的第3栏。模型 *M*2 与 *M*1 的对数似然值均显示在同一个表格中。似然比值统计量为0.96,3个自由度,尽管我们在此采用了"数据探测"的手段,然而该约束无副作用。

现在对 *M*1 增加下述约束,得 *M*3,即

$$\Theta = I$$

也可对 *M*2 设定 $\theta_{44}=1$ 而得到。比较模型 *M*3 与 *M*2,其对数比值统计量为58.52,1个自由度,从而可以拒绝模型 *M*3。对于题器[Friend],其残方差明显小于其他几个题器,而这几个题器的残方差相等。

**表10.11 模型 *M*1 与 *M*2 的参数估计**

| | *M*1 | *M*2 |
|---|---|---|
| 固定部分 | | |
| 阈值 | | |
| $\kappa_1$ | 0 | 0 |
| $\kappa_2$ | 1.08(0.04) | 1.10(0.02) |
| $\kappa_3$ | 1.64(0.05) | 1.67(0.03) |
| $\kappa_4$ | 2.25(0.07) | 2.29(0.03) |
| $\kappa_5$ | 2.64(0.08) | 2.69(0.04) |
| $\kappa_6$ | 3.13(0.09) | 3.19(0.05) |
| 截距 | | |
| $\beta_1$[City] | 1.13(0.05) | 1.15(0.04) |
| $\beta_2$[Hobby] | 0.82(0.05) | 0.84(0.04) |
| $\beta_3$[Family] | 0.22(0.05) | 0.22(0.05) |
| $\beta_4$[Friend] | 0.44(0.04) | 0.45(0.04) |
| $\beta_5$[Health] | 0.81(0.05) | 0.83(0.04) |

续表

| | $M1$ | $M2$ |
|---|---|---|
| 随机部分 | | |
| 因子负荷 | | |
| $\lambda_1$[City] | 1 | 1 |
| $\lambda_2$[Hobby] | 1.44(0.18) | 1.43(0.17) |
| $\lambda_3$[Family] | 1.98(0.25) | 1.96(0.23) |
| $\lambda_4$[Friend] | 1.81(0.22) | 1.82(0.19) |
| $\lambda_5$[Health] | 1.44(0.18) | 1.43(0.17) |
| 因子方差 | | |
| $\psi_{11}$ | 0.48(0.04) | 0.49(0.04) |
| 残方差 | | |
| $\theta_{11}$[City] | 1 | 1 |
| $\theta_{22}$[Hobby] | 0.97(0.09) | 1 |
| $\theta_{33}$[Family] | 0.93(0.09) | 1 |
| $\theta_{44}$[Friend] | 0.45(0.05) | 0.47(0.05) |
| $\theta_{55}$[Health] | 0.93(0.08) | 1 |
| 对数似然值 | -7 669.26 | -7 669.75 |

接下来，对于生活满意度题器，让我们探讨一个可能的二维模型。由于对这些题器因子负荷的模式没有预判，我们设定一个弱二维性探索性因子模型。为了模型辨识，设定：

$$\lambda_{21} = 1 = \lambda_{42}$$

与

$$\lambda_{11} = 0 = \lambda_{52}$$

式(10.2)的约束仍然保留。该模型被标记为 $M0$，辨识约束为 $\kappa_1 = 0$ 并且 $\theta_{11} = 1$。请注意，单维模型 $M2$ 嵌套于二维模型 $M0$ 之中。模型 $M2$ 与 $M0$ 的似然值非常接近，我们认为该生活满意度的单维模型的约束是可以接受的，因而保留该模型。通过如图10.7所示的二维生活满意度模型的经验贝叶斯因子得分进一步支持了这个结论。请注意模

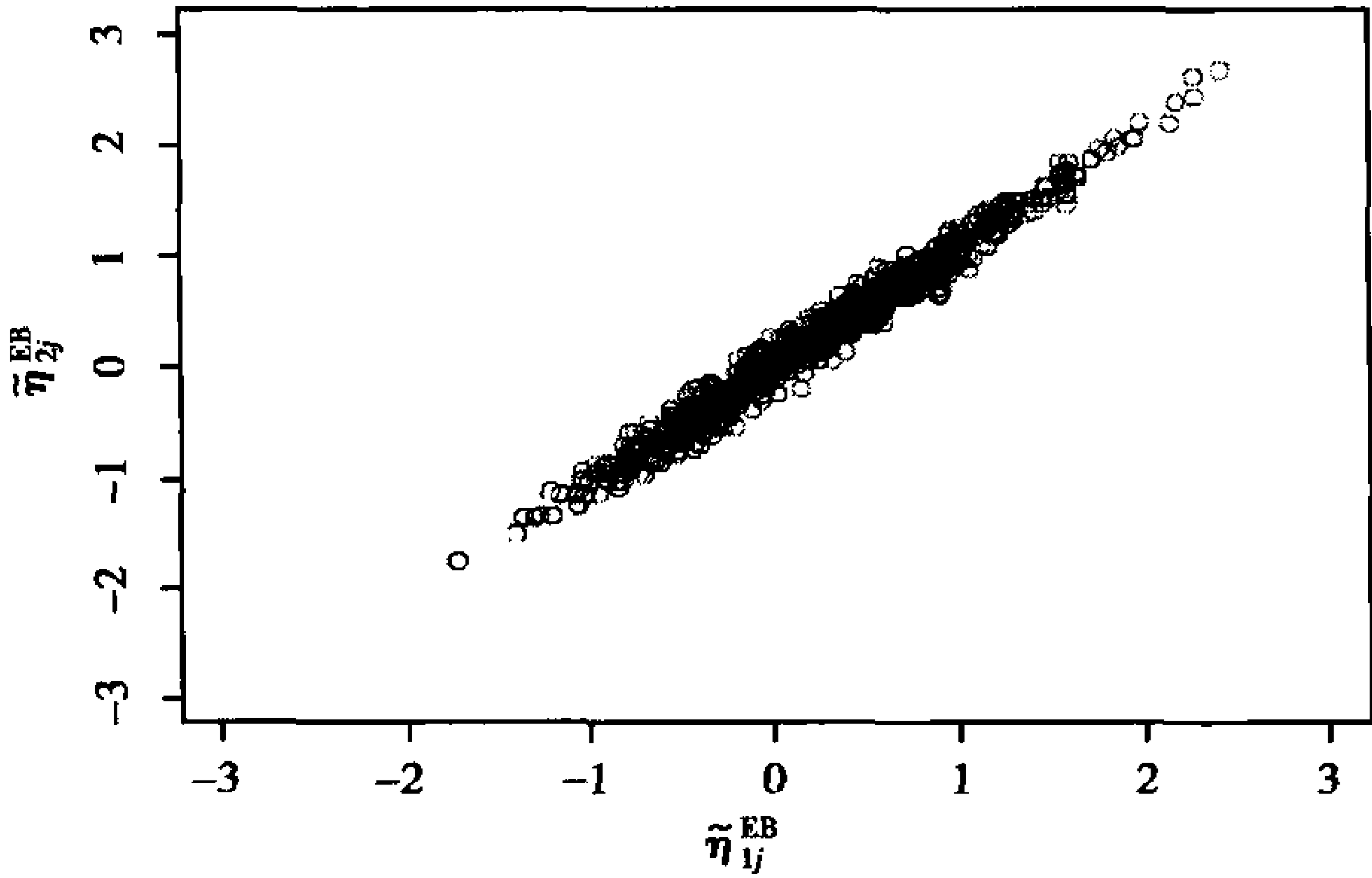



**图 10.7　生活满意度的经验贝叶斯因子得分($M4$)**

型 *M*0 的得分接近于线性关联,同样也反映在所估计的相关系数$\frac{\hat{\psi}_{21}}{\sqrt{(\hat{\psi}_{11}\hat{\psi}_{22})}}=0.93$ 上。对于探索性因子模型来说,这个结果是相当惊人的。

可见,模型 *M*2 优于其他模型,因而作为我们的保留模型。根据拉扎斯菲尔德(Lazarsfeld,1950),我们将模型 *M*2 的**题器特质曲线**如图 10.8 和图 10.9 所示。题器特质曲线表示条件响应分布(见 3.3.4 节),可通过下式得

$$\Pr(y_{ij}=s\mid\eta_j)=\Phi\left(\frac{\kappa_s-\nu_{ij}}{\sqrt{\theta_{ii}}}\right)-\Phi\left(\frac{\kappa_{s-1}-\nu_{ij}}{\sqrt{\theta_{ii}}}\right),\qquad s=1,\cdots,7$$

其中,$\kappa_0=-\infty$,$\kappa_7=\infty$。题器特质曲线将对于给定题器 $i$,响应在其中某个类别 $s$ 上的概率呈现为关于生活满意度程度的函数(在图中,我们已逆转了生活满意度量表的方向,使得高满意度与低响应类别相关联)。从图 10.8 可以明显看出,题器[Friend]比题器[City]更好,因为在不同的满意程度下,响应类别得到很好的区分。由图 10.9 可知,其他题器在这方面处于中等程度。

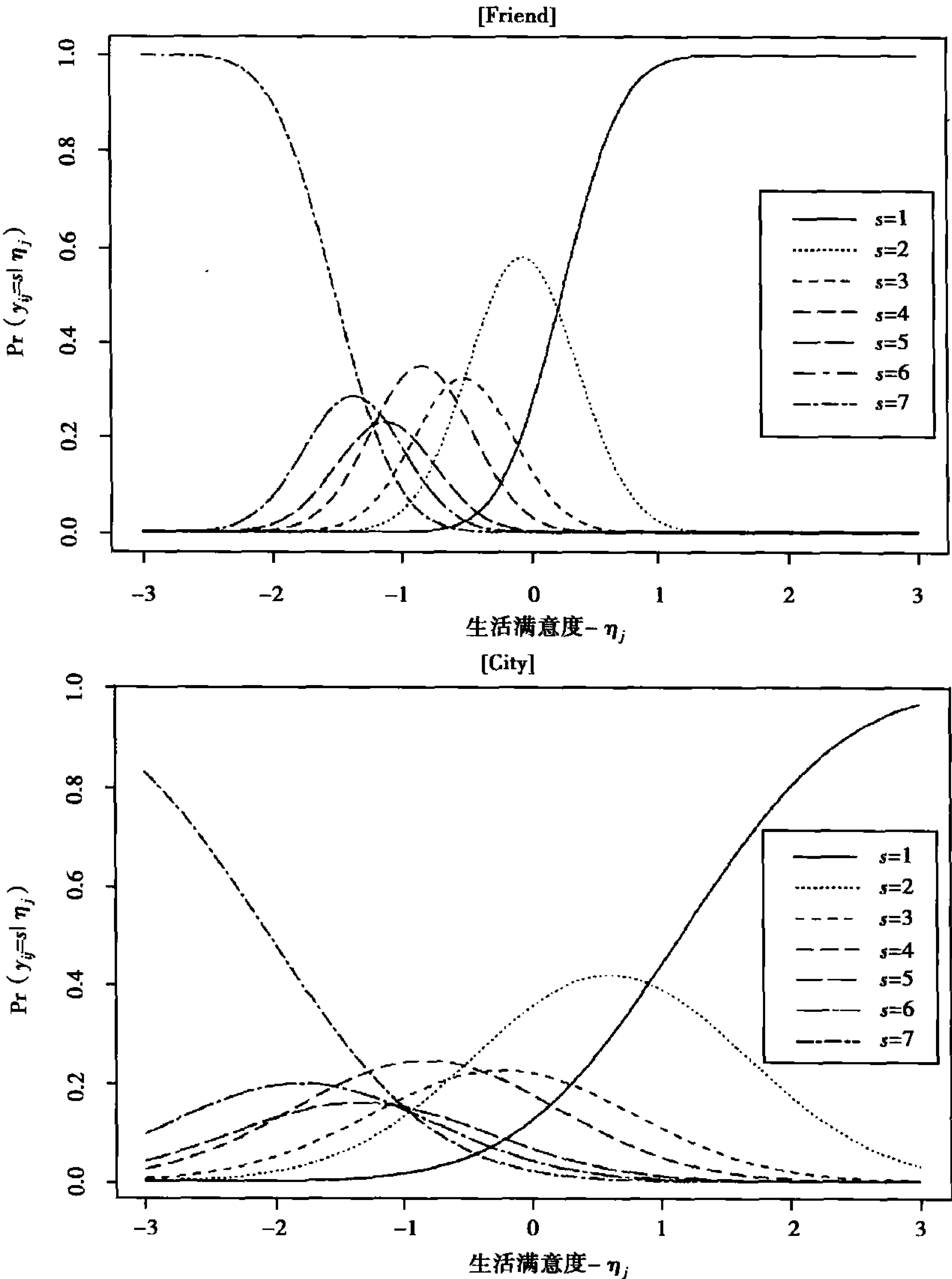



**图 10.8 题器[Friend]与[City]的条件响应分布(*M*2)**

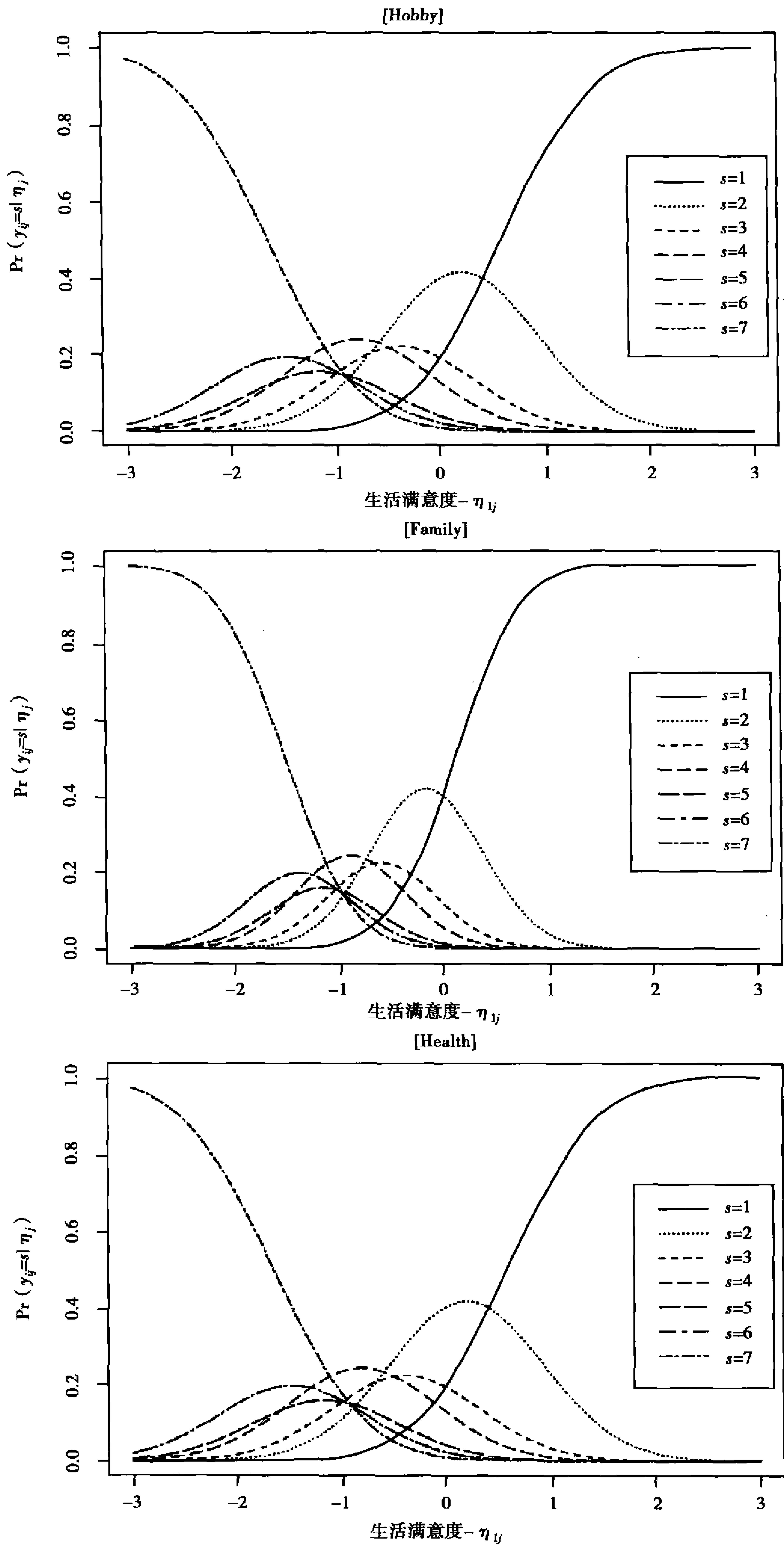



图 10.9 题器[Hobby],[Family]与[Health]的条件响应分布($M2$)

### 10.4.3 信　度

接下来，在这个单维模型中，我们将分析这些生活满意度题器的信度。与 10.3.2 节所介绍的定义类似，我们将该模型的信度 $\rho_i$ 下限定义为

$$\rho_i = \frac{\mathbf{\Lambda}_i \psi_{11} \mathbf{\Lambda}_i'}{\mathbf{\Lambda}_i \psi_{11} \mathbf{\Lambda}_i' + \theta_{ii}}$$

模型 *M*1 与 *M*2 所估计的信度下限如表 10.12 所示的第 2 栏和第 3 栏。

**表 10.12　模型 *M*1，*M*2 与 *M*4 所估计的信度下限**

| | *M*1 | *M*2 | *M*4 |
|---|---|---|---|
| [City] | 0.19 | 0.19 | 0.21 |
| [Hobby] | 0.33 | 0.33 | 0.35 |
| [Family] | 0.49 | 0.48 | $0.51^{=}$ |
| [Friend] | 0.62 | 0.63 | $0.51^{=}$ |
| [Health] | 0.34 | 0.33 | 0.35 |

我们可知，题器[City]的信度下限非常低，题器[Hobby]与[Health]的信度下限处于中间位置，而题器[Family]与[Friend]的信度下限远远超出其他题器。题器[City]的效用很值得怀疑，因为它的测量误差或**题器成分**(item specific Component)相当大。

接下来的问题则是，对于这两个信度下限很高的题器[Family]与[Friend]，它们的下限是否相等？为了分析这个问题，正好借此机会展示一下在 4.5 节所讨论的标量参数约束策略(scalar parameter restriction strategy)的效力。我们希望增加下面的约束：

$$\rho_3 = \rho_4$$

将 $\rho_i$ 的定义替代入公式，并且在模型 *M*2 中 $\theta_{33} = 1$，可得

$$\frac{(\lambda_3)^2 \psi_{11}}{(\lambda_3)^2 \psi_{11} + 1} = \frac{(\lambda_4)^2 \psi_{11}}{(\lambda_4)^2 \psi_{11} + \theta_{44}}$$

对 $\theta_{44}$解上述方程得

$$\theta_{44} = \left(\frac{\lambda_4}{\lambda_3}\right)^2$$

很显然，这是**非线性**参数约束。请注意，在这个例子中，$\theta_{44}$被当做结构参数 $\lambda_3$ 与 $\lambda_4$ 的函数，因而 $\theta_{44}$肯定无法用基本参数来表示。这个最终得到的模型包含非线性参数约束，标记为 *M*4。该模型结构参数与基本参数的转换表如表 10.13 所示。请特别注意 $\theta_{44}$ 的表达式。

至于所估计的参数，我们得到 $\hat{\lambda}_3 = 1.95$ 以及 $\hat{\lambda}_4 = 1.67$，因而 $\hat{\theta}_{44} = 0.73$。如果需要估计 $\hat{\theta}_{44}$的标准误，可通过 8.3.1 节介绍的德尔塔方法得到。信度下限估计值如表 10.7 所示的第 3 列，上标“=”表示存在相等约束。

比较嵌套模型 *M*4 与 *M*2，似然比值统计量为 18.89，1 个自由度，因而有足够的证据拒绝题器[Family]与题器[Friend]具有相等的信度下限。

表 10.13 结构参数与基本参数转换表(*M*4)

| 结构参数 | 基本参数 |
|---|---|
| $\kappa_{12},\kappa_{22},\kappa_{32},\kappa_{42},\kappa_{52}$ | $\theta_1$ |
| $\kappa_{13},\kappa_{23},\kappa_{33},\kappa_{43},\kappa_{53}$ | $\theta_2$ |
| $\kappa_{14},\kappa_{24},\kappa_{34},\kappa_{44},\kappa_{54}$ | $\theta_3$ |
| $\kappa_{15},\kappa_{25},\kappa_{35},\kappa_{45},\kappa_{55}$ | $\theta_4$ |
| $\kappa_{16},\kappa_{26},\kappa_{36},\kappa_{46},\kappa_{56}$ | $\theta_5$ |
| $\mu_1$ | $\theta_6$ |
| $\mu_2$ | $\theta_7$ |
| $\mu_3$ | $\theta_8$ |
| $\mu_4$ | $\theta_9$ |
| $\mu_5$ | $\theta_{10}$ |
| $\lambda_{21}$ | $\theta_{11}$ |
| $\lambda_{31}$ | $\theta_{12}$ |
| $\lambda_{41}$ | $\theta_{13}$ |
| $\lambda_{51}$ | $\theta_{14}$ |
| $\psi_{11}$ | $\theta_{15}$ |
| $\theta_{44}$ | $\theta_{13}/\theta_{12}$ |

### 10.4.4 特殊情形

穆拉齐(Muraki, 1990)提供了一个 IRT 模型,其实是本节所讨论模型的一个特例。顺便一提,我们注意到有很多特殊的定类因子模型与 IRT 文献中阐述的模型等价(如 Takane and de Leeuw, 1987)。使用本书附带的由穆拉齐和博科(Muraki and Bock, 1993)所写的软件 PARSCALE,可得到对该模型的 EM 算法估计。

首先,我们看到穆拉齐-博科模型只包含了单维模型。另外,除了和我们一样的辨识约束外,他们还设定 $\mathbf{\Theta}=\mathbf{I}$。并且,当因子负荷并不都相同时,设定 $\kappa_{s-2}=a$,其中 $a$ 为固定值。穆拉齐与博科认为,为了模型辨识这两个约束是必要的。然而,根据我们在 5.2.3 节的分析,很显然这些约束不是必要的。他们之所以没有认识到这一点,很显然是由于过于关注题器的阈值结构(见 Muraki, 1990:64),因而在相关结构及题器的**联立阈值结构**(simultaneous threshold structure)上所包含的信息没有得到利用。另外,他们的有些辨识约束在经验上是错误的。例如,如果我们比较模型 *M*2 与 *M*3,就可拒绝约束 $\mathbf{\Theta}=\mathbf{I}$。类似的,与二维的 *M*2 模型相比较,我们更倾向于保留单维 *M*2 模型。最后,穆拉齐-博科方法还存有一个缺陷,即无法包含协变量。在 5.2 节所概述的辨识分析具有一个重要的优点,一些传统上被认为是无法辨识的模型如果用它去做的话,可能是可以辨识的。

### 10.4.5 结 论

我们认为 GSS 中所测量的生活满意度应为单维现象。在具体题器的性质上,题器[Friend]在判别力及信度上都最佳。题器[City]与之相反,处于反方向上,其测量误差或

题器成分较高。从而,应该考虑放弃这个题器。

## 10.5 总结与进一步阅读

性别教育的例子涉及群组随机试验的复杂研究设计,关于潜变量的多元定序测量及重复测量。第一个分析的模型为随机截距比例比率模型。其他关于群组定序数据的模型请见阿奎斯提和纳塔拉杨(Agresti and Natarajan, 2001)。沃尔夫和费斯(Wolfe and Firth, 2002)允许阈值随机变化,而非设定一个随机截距。金等(King et al., 2003)提出一个方法,他们用晕影法将阈值锚定,从而可在不同文化之间比较有序响应。

接着,我们使用了潜增长曲线模型来对"避孕效能"建模。当潜变量的测量为连续变量(如 Rabe-Hesketh et al., 2001d),或者为混合变量时(如 Gueorguieva and Sanacora, 2003),仍然能使用该模型。斯科隆多等(Skrondal et al., 2002)将这个性别教育潜增长曲线模型进行扩展,使其能适用于**不可忽略的退出**(nonignorable dropout)的情形。

政治效能以及生活满意度例子用于展示使用潜变量模型来进行定序题器量表的心理测量校验。我们研究了在什么情况下可以将量表称为单维或多维。题器偏误的分析使用广义 MIMIC 模型。在生活满意度例子中,题器有 7 个类别。由于定序变量的类别增加会导致阈值参数的数量迅速增加,我们将所有题器的阈值设为相等,但允许潜变量的截距与残方差不同,从而得到一个有衡定序正态概率模型。约翰逊和阿尔伯特(Johnson and Albert, 1999, 第 6 章和第 7 章)及穆斯塔奇(Moustaki, 2000)讨论了定序 IRT 模型(或因子模型)。格里利和拉姆皮奇尼(Grilli and Rampichini, 2003)提出二元多层次定序响应模型(Bivariate multilevel ordinal response model)。

在本章,我们对于定序响应变量使用了累积模型(比例比率模型及定序正态概率模型)。其他可用的模型还包括临近类别对数概率模型或**连续比例对数概率**模型(continuation ratio logit)。后者在 12.3 节中用于对离散时间存续数据建模。我们没有讨论定序响应变量的潜类模型,可进一步参见沃芒(Vermunt, 2001)。

# 11 事件数

## 11.1 介 绍

在本节,我们讨论事件数变量的泊松模型及二项模型。由于事件数变量为聚合的数据,即通过对表示事件发生或不发生的二项数据进行汇总而得到,因此即使是非常简单的数据结构,如每个单元只有一个事件数的情形,也可当做双层数据集。对于事件数变量,可能存在未观测到的异质性,从而导致过度离散。

在本章的第一个例子中,我们讨论处理过度离散的几种方法,包括零堆积泊松(zero-inflated Poisson)模型和零堆积二项(zero-inflated binomial)模型。在第 2 个例子里,我们对纵贯性事件数数据估计随机系数模型,并使用 8.6 节所讨论的几种模型诊断方法。最后,我们讨论疾病制图(disease mapping)与小域估计(small area estimation),并使用带有空间依赖结构的模型来估计随机效应。

## 11.2 儿童牙病防治:过度离散模型

### 11.2.1 介 绍

我们将分析一个关于巴西儿童牙齿健康的数据,共有 797 个儿童参加了牙齿健康诊疗(Mendonca and Böhning, 1994)[1]。共有 6 所学校,每所学校被分配 6 种预防龋齿诊疗方法的其中一种。

- [Control] 没有参加试验。
- [Educ] 口头健康教育。
- [Enrich] 学校提供富含米糠的食品。
- [Rinse] 含 0.2% 浓度氟化钠混合液的固龈液。
- [Hygiene] 口腔保健。
- [All] 上述全部 4 种试验方法。

结果变量为龋齿、缺齿或补牙(DMFT)的个数。除了学校以外,另外还有两个协变量:

- [Male] 男孩,虚拟变量。

---

1 该数据可从 gllamm. org/books 或皇家统计学会(Royal Statistical Society)数据库网站 http://www. blackwellpublishing. com/rss/Volumes/av162p2. htm 下载。

● 种族:(参照组为“棕色人种”)。

——[White] 白人,虚拟变量。

——[Black] 黑人,虚拟变量。

观测到的 DMFT 事件数(边缘对协变量)的分布显示在表 11.1 的前两列中。显然应首先考虑 2.2 节中介绍的泊松模型,对于儿童 $i$,其均值为 $\mu_i$,建模如下

$$\log(\mu_i) = \mathbf{x}'_i\boldsymbol{\beta}$$

其中,协变量 $\mathbf{x}_i$ 为试验组、性别及种族的虚拟变量。表 11.1 第 3 列显示了该模型所估计的频次,同样的,也是边缘于协变量。频次估计通过如下计算方法获得:对于每一个观测到的协变量组合,计算事件数从 0 到 20 之间的概率,乘上样本总体的概率,再对协变量加总。尽管所有可能的协变量都已包含在模型中,在观测值与期望值之间仍存在较大的差异。这个差异主要来自事件数为 0 的 DMFT(231 个观测值对比 134 个期望值)。

**表 11.1 不同类型的过度离散泊松模型的观测频次与期望频次**

**(关于协变量的边缘和随机效应)**

| DMFT 事件数 | 观测频次 | 期望频次 | | | |
|---|---|---|---|---|---|
| | | 泊松 | 正态截距 | 双分类截距 | ZIP |
| 0 | 231 | 133.96 | 186.35 | 227.19 | 227.04 |
| 1 | 163 | 229.64 | 221.64 | 168.24 | 148.44 |
| 2 | 140 | 205.75 | 165.50 | 147.66 | 165.05 |
| 3 | 116 | 128.14 | 101.72 | 115.94 | 125.70 |
| 4 | 70 | 62.21 | 56.95 | 73.12 | 73.63 |
| 5 | 55 | 25.03 | 30.51 | 38.06 | 35.31 |
| 6 | 22 | 8.66 | 16.06 | 16.92 | 14.42 |
| 7 | 0 | 2.64 | 8.44 | 6.59 | 5.15 |
| 8 | 0 | 0.72 | 4.46 | 2.29 | 1.64 |
| 9 | 0 | 0.18 | 2.39 | 0.72 | 0.47 |
| 10 | 0 | 0.04 | 1.29 | 0.21 | 0.12 |
| 11 | 0 | 0.01 | 0.71 | 0.06 | 0.03 |
| 12 | 0 | 0.00 | 0.40 | 0.01 | 0.01 |
| 13 | 0 | 0.00 | 0.23 | 0.00 | 0.00 |
| 14 | 0 | 0.00 | 0.13 | 0.00 | 0.00 |
| 15 | 0 | 0.00 | 0.08 | 0.00 | 0.00 |
| 16 | 0 | 0.00 | 0.05 | 0.00 | 0.00 |
| 17 | 0 | 0.00 | 0.03 | 0.00 | 0.00 |
| 18 | 0 | 0.00 | 0.02 | 0.00 | 0.00 |
| 19 | 0 | 0.00 | 0.01 | 0.00 | 0.00 |
| 20 | 0 | 0.00 | 0.01 | 0.00 | 0.00 |

### 11.2.2 对过度离散建模

由于有大量的观测零值,波宁等(Böhning et al., 1999)估计了一个**零堆积泊松模型**(ZIP)(Lambert, 1992)。ZIP 模型为两个泊松分布的组合,其中一个分布的均值为 0,另一个分布的均值依赖于协变量,其响应概率为

$$\Pr(y_i \mid \mathbf{x}_i) = \pi_1 g(y_i;\mu_i = 0) + \pi_2 g(y_i;\mu_i = \exp(\mathbf{x}'_i\boldsymbol{\beta})) \qquad (11.1)$$

其中,$\pi_1$ 和 $\pi_2 = 1 - \pi_1$ 为成分权重(component weight)或称潜类概率,$g(y_i;\mu_i)$是均值为 $\mu_i$ 的事件数 $y_i$ 的泊松概率

$$g(y_i;\mu_i) = \frac{\mu_i^{y_i}\exp(-\mu_i)}{y_i!}$$

当 $y_i > 0$,式(11.1)的第 1 项为 0,使得事件数大于 0 的单元属于类别 2。然而,当 $y_i = 0$ 时,公式的两项均大于 0,因为零事件数即可能来自均值为 0 的泊松模型(类别 1),也可能来自均值大于 0 的泊松模型(类别 2)。因而,事件数为 0 的单元都有可能属于其中一个类别。例如,如果事件数为过去两周所消费的饮酒数,0 响应既可能来自从不喝酒的人,也可能来自一个平时喝酒但是在这个期间正好没有喝酒的人。与普通泊松模型不同,在 ZIP 模型中,潜类概率 $\pi_1$ 确定零值的数量。

然而,这种显式允许大量零值的方式可能没有必要,因为其他形式的过度离散同样也呈现为大量零值。因而我们考虑使用随机截距模型

$$\ln(\mu_i) = \mathbf{x}'_i\boldsymbol{\beta} + \zeta_i$$

其中,$\zeta_j$ 符合 $N(0, \psi)$ 分布或概率为 $\pi_c$ 的离散分布

$$\zeta_i = e_c, \quad c = 1,\cdots,C$$

在这个例子中,当 $C = 2$ 时似然值最大,因而与非参最大似然估计量(NPMLE)相对应。值得注意的是,截距 $\zeta_i$ 在第一层次上变化,与本书的其他例子不同,在其他例子中它们均在更高的层次上变化。

传统的泊松模型、双分类 NPMLE 模型,正态随机截距泊松模型以及 ZIP 模型的估计结果显示在表 11.2 中,其对应的频次期望值显示在表 11.1 中。分别在 8.4.2 节与8.4.3 节所定义的 BIC 与 AIC 值显示在表 11.2 中,$N$ 值为儿童个数(797),$v$ 值为所估计的参数个数(正如 8.4.2 节所讨论的,判断 AIC 与 BIC 的 $N$ 与 $v$ 值在潜变量模型中并不清楚)。在表 11.2 中,最小的 AIC 与 BIC 值用粗体显示。

根据 AIC 统计量,双分类随机截距模型提供了最佳的拟合,但是根据 BIC 统计量,ZIP 模型则提供了最佳拟合。即使是拟合得最好的模型,观测值与期望值的差异在DMFT 事件数为 5 时特别大。

表 11.2 中的参数估计表示所估计的 DMFT 期望值的修正对数比值(adjusted log ratio)。例如,对于 ZIP 模型,接受全部试验方法组[All]的期望事件数除以控制组的期望事件数后再按性别与种族修正后的比值为 $\exp(-0.47) = 0.63$。其他试验方法也能削减 DMFT 期望值,但是[All]最有效果,[Enrich]只有可忽略不计的效果。比较普通泊松模型与正态随机截距泊松模型,我们可以看到协变量的效应几乎相同。这是因为对使用对数链接函数的随机截距模型来说,条件效应等于边缘效应,见 4.8.1 节。

在这个例子中使用泊松分布有点不恰当,因为表示龋齿、缺齿或补牙的事件数("成功")是按总数固定的 8 颗乳磨牙("试验")来计算的。因而,我们对这个事件数数据使用分母为 8 的二项分布模型。表 11.3 显示简单二项对数概率回归模型并不能产生足够多的 0 事件数期望频次。

表 11.2 不同泊松模型估计值

| 参数 | 泊松 | | 正态截距 | | 双分类截距 | | ZIP | |
|---|---|---|---|---|---|---|---|---|
| | Est | (SE) | Est | (SE) | Est | (SE) | Est | (SE) |
| 回归系数 | | | | | | | | |
| $\beta_0$[Cons] | 0.76 | (0.07) | 0.63 | (0.09) | — | | 0.94 | (0.08) |
| 试验: | | | | | | | | |
| $\beta_1$[Educ] | -0.23 | (0.09) | -0.23 | (0.11) | -0.24 | (0.11) | -0.22 | (0.09) |
| $\beta_2$[Enrich] | -0.09 | (0.09) | -0.09 | (0.11) | -0.08 | (0.09) | -0.06 | (0.09) |
| $\beta_3$[Rinse] | -0.35 | (0.08) | -0.37 | (0.11) | -0.26 | (0.10) | -0.22 | (0.09) |
| $\beta_4$[Hygiene] | -0.30 | (0.09) | -0.32 | (0.11) | -0.22 | (0.11) | -0.23 | (0.10) |
| $\beta_5$[All] | -0.59 | (0.10) | -0.61 | (0.12) | -0.49 | (0.11) | -0.47 | (0.11) |
| 性别: | | | | | | | | |
| $\beta_6$[Male] | 0.13 | (0.05) | 0.13 | (0.07) | 0.10 | (0.06) | 0.10 | (0.06) |
| 种族: | | | | | | | | |
| $\beta_7$[White] | 0.09 | (0.06) | 0.10 | (0.07) | 0.09 | (0.07) | 0.08 | (0.06) |
| $\beta_8$[Black] | -0.14 | (0.09) | -0.16 | (0.11) | -0.12 | (0.10) | -0.12 | (0.10) |
| 方差 | | | | | | | | |
| $\psi$ | — | | 0.29 | (0.05) | — | | — | |
| Log odds 参数 | | | | | | | | |
| $\varrho_0$ | — | | — | | -0.86 | (0.21) | -1.39 | (0.12) |
| 原点参数 | | | | | | | | |
| $e_1$ | — | | — | | -1.11 | (0.37) | $-\infty$ | — |
| $e_2$ | — | | — | | 1.04 | (0.09) | — | |
| 对数似然值 | -1 469.05 | | -1 432.53 | | -1 406.03 | | -1 410.27 | |
| AIC | 2 956.04 | | 2 885.07 | | **2 834.06** | | 2 840.54 | |
| BIC | 3 058.35 | | 2 998.68 | | 2 959.04 | | **2 954.16** | |

为了处理这个问题，与前面讨论的 ZIP 模型类似，我们引入**零堆积二项模型**(zero-inflated binomial (ZIB))。ZIB 模型为两个二项分布的组合，其中一个概率参数等于零，另一个概率参数通过对数概率链接函数依赖于协变量，其响应概率为

$$\Pr(y_i \mid \mathbf{x}_i) = \pi_1 g(y_i;\mu_i = 0) + \pi_2 g(y_i;\text{logit}(\mu_i) = \mathbf{x}'_i\boldsymbol{\beta}) \qquad (11.2)$$

其中，$g(y_i;\mu_i)$是参数为$\mu_i$，分母为 8 的二项概率，则

$$g(y_i;\mu_i) = \frac{y!}{y!(8-y)!}\mu_i^{y_i}(1-\mu_i)^{8-y_i}$$

同样的，我们可以估计包含正态随机截距或非参随机截距的二项对数概率回归模型。该模型的估计结果显示在表 11.4 中，期望频次显示在表 11.3 中。

表 11.3 不同类型的过度离散二项对数概率回归模型的观测频次和预测频次关于协变量的边缘和随机效应

| DMFT 事件数 | 观测频次 | 预测频次 | | | |
|---|---|---|---|---|---|
| | | 二项 | 正态截距 | 三分类截距 | ZIB |
| 0 | 231 | 107.21 | 207.22 | 226.73 | 227.88 |
| 1 | 163 | 230.82 | 202.14 | 170.45 | 120.50 |
| 2 | 140 | 233.84 | 149.66 | 143.67 | 174.21 |
| 3 | 116 | 145.04 | 101.13 | 108.65 | 150.21 |
| 4 | 70 | 59.90 | 64.65 | 78.02 | 84.21 |
| 5 | 55 | 16.76 | 38.81 | 45.52 | 31.33 |
| 6 | 22 | 3.08 | 21.09 | 18.70 | 7.53 |
| 7 | 0 | 0.34 | 9.51 | 4.71 | 1.06 |
| 8 | 0 | 0.02 | 2.78 | 0.55 | 0.07 |

表 11.4 各种二项对数概率回归模型估计值

| 参数 | 二项 | 正态截距 | 三分类截距 | ZIB |
|---|---|---|---|---|
| | Est (SE) | Est (SE) | Est (SE) | Est (SE) |
| 回归系数 | | | | |
| $\beta_0$[Cons] | -1.00(0.08) | -1.29(0.14) | — | -0.69(0.09) |
| 控制组 | | | | |
| $\beta_1$[Educ] | -0.32(0.10) | -0.35(0.17) | -0.41(0.17) | -0.32(0.11) |
| $\beta_2$[Enrich] | -0.12(0.10) | -0.15(0.17) | -0.16(0.16) | -0.09(0.11) |
| $\beta_3$[Rinse] | -0.47(0.10) | -0.57(0.16) | -0.49(0.16) | -0.26(0.11) |
| $\beta_4$[Hygiene] | -0.40(0.10) | -0.50(0.18) | -0.36(0.17) | -0.29(0.11) |
| $\beta_5$[All] | -0.76(0.11) | -0.91(0.17) | -0.80(0.17) | -0.59(0.13) |
| 性别: | | | | |
| $\beta_6$[Male] | 0.17(0.06) | 0.20(0.10) | 0.16(0.09) | 0.13(0.07) |
| 种族: | | | | |
| $\beta_7$[White] | 0.13(0.07) | 0.14(0.11) | 0.13(0.10) | 0.11(0.08) |
| $\beta_8$[Black] | -0.18(0.10) | -0.23(0.16) | -0.17(0.15) | -0.14(0.12) |
| 方差 | | | | |
| $\psi$ | — | 1.05(0.12) | — | — |
| Log odds 参数 | | | | |
| $\varrho_0^1$ | — | — | -1.01(0.40) | -1.16(0.10) |
| $\varrho_0^2$ | — | — | 0.27(0.30) | — |
| 原点参数 | | | | |
| $e_1$ | — | — | -32.09* | -∞ — |
| $e_2$ | — | — | -1.53(0.26) | — |
| $e_3$ | — | — | -0.06(0.18) | — |
| 对数似然值 | -1 546.78 | -1 409.39 | -1 397.53 | -1 431.09 |
| AIC | 3 111.56 | 2 838.78 | 2 821.07 | 2 882.19 |
| BIC | 2 113.81 | 2 952.40 | 2 968.77 | 2 995.80 |

注：* 为边界解。

值得一提的是,二项对数概率正态随机截距模型拟合得比泊松正态随机截距模型要好些(在参数个数相同的情况下,对数似然值 -1 409.39 对 -1 432.53),这主要是因为后者根据高的事件数值得到相当高的期望频次,而前者无法得到超过 8 的期望频次。二项对数概率 NPML 方案具有 3 个类别(或质量),根据 AIC,这个模型在所有考虑的模型中拟合得最好;但是如果根据 BIC,则是正态随机截距二项对数概率模型拟合得最好。

二项对数概率模型的参数估计值表示对数优比估计值。例如,对于三分类模型,根据性别与种族所修正的[All]组儿童具有一个缺陷牙齿(DMFT)的比率除以控制组儿童具有一个缺陷牙齿(DMFT)的比率的比值为 $\exp(-0.80)=0.45$。同样的,所有的试验手段都有一定的收益,尽管[Enrich]的效应看上去可以忽略不计。

随机截距二项模型表示给定随机截距的条件效应,普通二项模型表示边缘效应,比较这两个模型,可以看到 4.8.1 节所讨论的衰减是存在的。请注意,由于第一个类别所估计的原点为 -32.09,因而二项 NPML 模型的估计值表示边界解,对应于几乎等于零的二项概率参数。属于这个类别的对数比率(log-odds)估计值为 -1.01,差不多等于 ZIB 模型中属于零概率类别的对数比率 -1.16(其对应的概率分别为 0.27 和 0.24)。

## 11.3 癫痫症治疗:随机系数模型

### 11.3.1 介 绍

莱皮克等(Leppik et al., 1987)的纵贯性癫痫症数据之前由 Thall and Vail (1990)、Breslow and Clayton (1993)、Lindsey (1999)、Diggle et al. (2002) 以及其他许多学者做过分析。该数据来自抗癫痫药的随机对照试验[1]。在 8 周的观察期内,每个病人癫痫发作的次数被记录下来。然后将病人随机分配到给予药品的试验组或给予安慰剂的控制组(含标准的化疗方法)。结果为在 4 次连续的诊疗之前的两周内癫痫发作的次数。布里斯罗与克雷顿使用了下述协变量:

- [Lbas] 在接受诊疗前 8 周内癫痫发作的次数除以 4 再求对数。
- [Treat] 是否属于试验组,虚拟变量。
- [LbasTrt] 上述两个变量的交互。
- [Lage] 年龄的对数。
- [V4] 第 4 次临诊,虚拟变量。
- [Visit] 临诊次数,编码为 -0.3, -0.1, 0.1 及 0.3。

### 11.3.2 重复事件数数据建模

布里斯罗与克雷顿的模型Ⅱ为对数线性模型(泊松回归),包含上面所列的除了[Visit]以外的全部协变量以及每个病人的截距。病人 $j$ 在临诊 $i$ 的癫痫发作次数 $y_{ij}$ 被假设为均值为 $\mu_{ij}$ 的条件泊松分布,其模型为

$$\nu_{ij} = \log(\mu_{ij}) = \mathbf{x}'_{ij}\boldsymbol{\beta} + \zeta_{1j}$$

其中,个体特定的随机截距 $\zeta_{1j}$ 被假定为均值为零,方差为 $\psi_{11}$ 的正态分布。

布里斯罗与克雷顿的模型Ⅳ用预测变量[Visit]替代[V4],除了随机截距之外还包

1 数据可从 gllamm.org/books 下载。

括[Visit]的随机斜率 $\zeta_{2j}$。

$$\nu_{ij} = \log(\mu_{ij}) = \mathbf{x}'_{ij}\boldsymbol{\beta} + \zeta_{1j} + \zeta_{2j}z_{ij}$$

截距与斜率被假设为服从方差分别为 $\psi_{11}$ 与 $\psi_{22}$，协方差为 $\psi_{21}$ 的二元正态分布。拉贝-赫斯基思等(Rabe-Hesketh et al., 2002) 证明这个模型的参数可以通过自适应积分法得到可靠的估计值,但用普通积分法会非常不稳定(见6.3.2节关于积分法的描述)。因而我们对于模型Ⅱ使用每个维度15个点的自适应积分法,对模型Ⅳ使用每个维度8个点的自适应积分法。最大似然估计值及标准差显示在表11.5中,另外,还包括布里斯罗与克雷顿所报告的PQL-1估计值(见6.3.1节)。用自适应积分法得到的估计结果与PQL-1非常类似(由于我们将预测变量按均值居中,因而无法比较常数项),进一步确信了PQL-1可用于这种特殊例子。

**表11.5 采用PQL-1 (Breslow and Clayton, 1993)以及自适应高斯积分最大似然估计的模型Ⅱ和Ⅳ的参数与标准误**

| | 模型Ⅱ | | 模型Ⅳ | |
|---|---|---|---|---|
| | PQL-1 | AGQ | PQL-1 | AGQ |
| | Est(SE) | Est(SE,$SE_R$)$^+$ | Est(SE) | Est(SE,$SE_R$)$^\dagger$ |
| 固定效应 | | | | |
| $\beta_0$[Cons] | -1.25(1.2) | 2.11(0.22,0.21) | -1.27(1.2) | 2.10(0.22,0.21) |
| $\beta_1$[Lbas] | 0.87(0.14) | 0.88(0.13,0.11) | 0.87(0.14) | 0.89(0.13,0.11) |
| $\beta_2$[Treat] | -0.91(0.41) | -0.93(0.40,0.40) | -0.91(0.41) | -0.93(0.39,0.40) |
| $\beta_3$[LbasTrt] | 0.33(0.21) | 0.34(0.20,0.20) | 0.33(0.21) | 0.34(0.20,0.20) |
| $\beta_4$[Lage] | 0.47(0.35) | 0.48(0.35,0.30) | 0.46(0.36) | 0.48(0.35,0.33) |
| $\beta_5$[V4] | -0.16(0.05) | -0.16(0.05,0.07) | | |
| $\beta_6$[Visit] | — | — | -0.26(0.16) | -0.27(0.16,0.17) |
| 随机效应 | | | | |
| $\sqrt{\psi_{11}}$ | 0.53(0.06) | 0.50(0.06,0.06) | 0.52(0.06) | 0.50(0.06,0.06) |
| $\sqrt{\psi_{22}}$ | — | — | 0.74(0.16) | 0.73(0.16,0.16) |
| $\psi_{21}$ | — | — | -0.01(0.03) | 0.00(0.09,0.11) |
| 对数似然值 | — | -665.29 | — | -655.68 |

注:† $SE_R$表示基于夹挤估计量的“稳健”标准误。

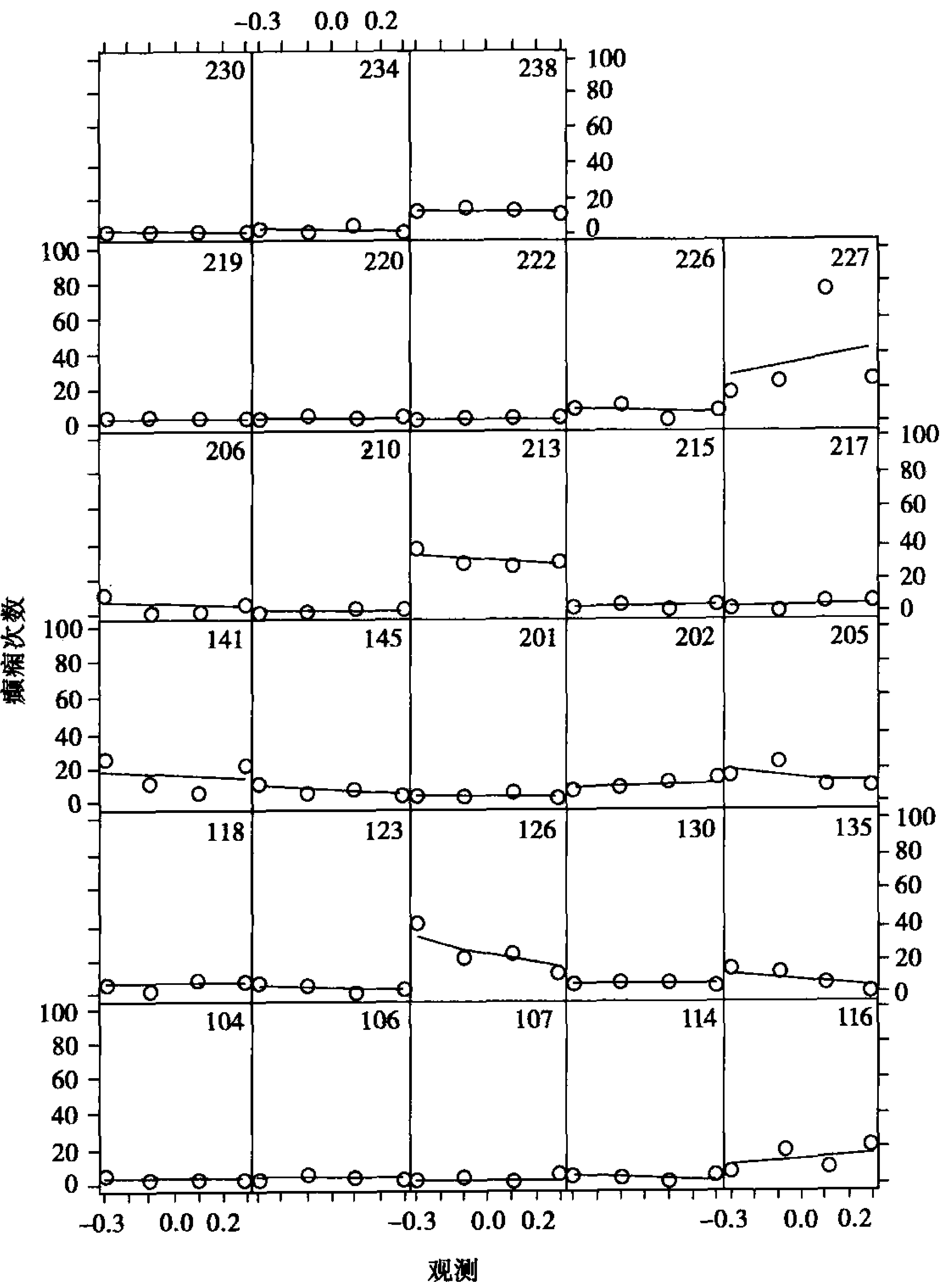
请注意,姚和库克(Yau and Kuk, 2002)所报告的根据16点普通积分法所得到的模型Ⅱ的估计结果与用PQL-1及自适应积分法的估计结果有相当的差异(如试验效应估计为-0.52),这表明他们的结果不可靠。对于最大似然估计,我们同时提供了基于夹挤估计量的“稳健”标准误 $SE_R$(见8.3.3节)。

图11.1显示了模型Ⅳ所估计的试验组中每个病人在每次临诊时癫痫次数的增长曲线,另外,观测频数用圆圈表示。在此,次数估计为线性估计量的指数幂的后验均值,即

$$E_{\zeta}[\exp(\mathbf{x}'_{ij}\boldsymbol{\beta} + \zeta_{1j} + \zeta_{2j}z_{ij}) \mid \mathbf{y}_j, \mathbf{X}_j, \mathbf{z}_j]$$

正如我们在7.8节所指出的,应该在考虑后验分布的情况下合并上面的指数方程以获得期望值,而非简单地将经验贝叶斯预测 $\tilde{\zeta}_{1j}$与 $\tilde{\zeta}_{2j}$代入上面的指数方程。

对于发作次数高的病人，斜率上的变差性非常明显。227个病人所观测到的次数与



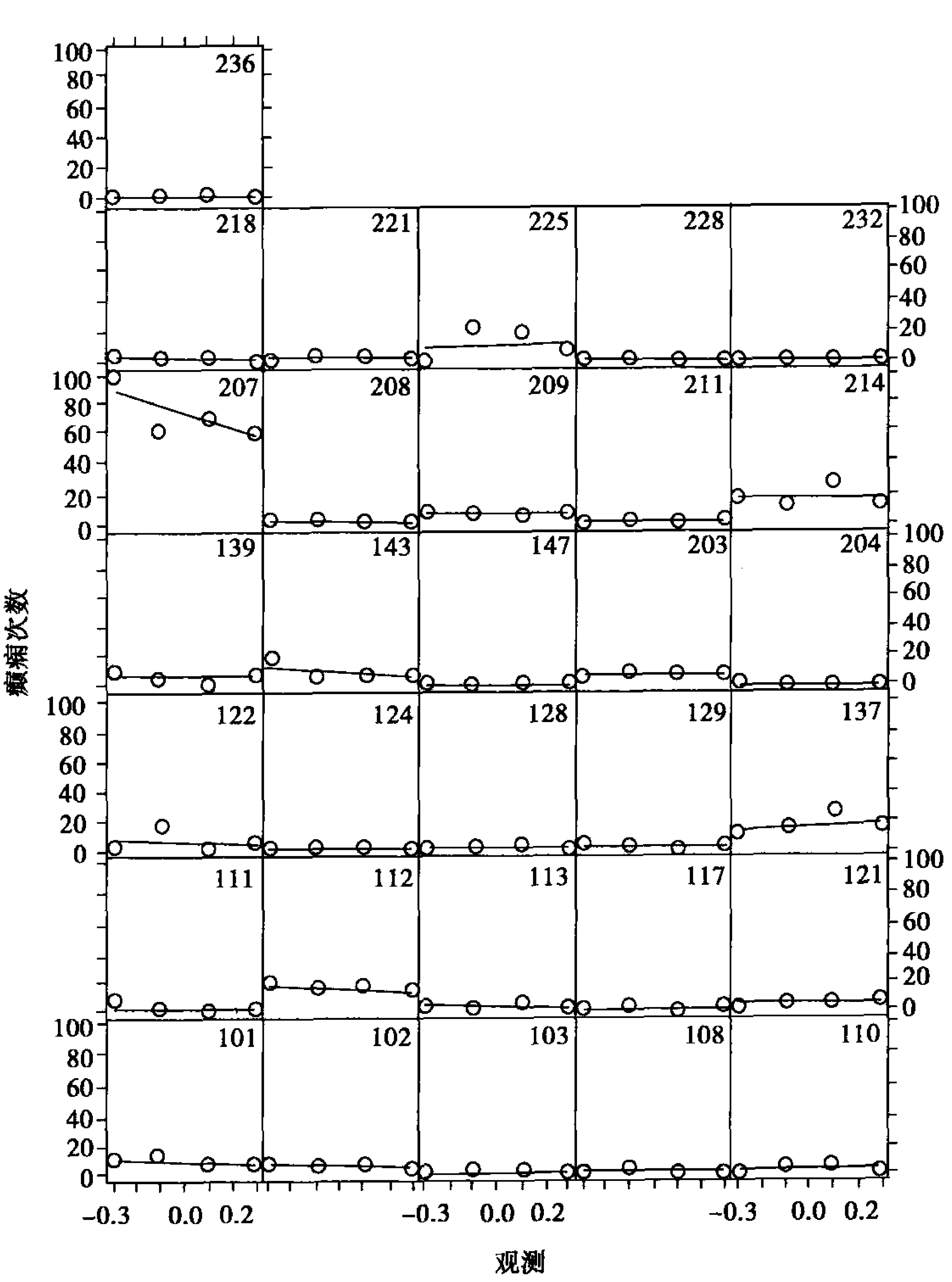




**图11.1 癫痫次数观测值和预测值与[Visit]**

预测次数有显著差异,特别是在第 3 次临诊的时候。

### 11.3.3　模型诊断

对只包含随机截距的模型Ⅱ,我们现在用 8.6 节讨论的模型诊断方法进行分析。

随机效应的正态性可以通过估计一个非参随机截距分布模型来进行评估。NPML 法(没有进行均值居中处理)有 6 个质量, -30, 1.00, 1.75, 2.03, 2.37 及 2.90,其概率分别为 0.02, 0.15, 0.45, 0.22, 0.15 及 0.09(姚与库克(Yau and Kuk, 2002)仅发现 4 个质量)。5 个质量的解为边界解,并且其中一个类别在期望值为 0 时,其原点是很大的负数,为了避免出现扁平的对数似然及获得 NPMLE,原点被设定为 -30。图11.2显示了当所有均值居中的协变量为零时所预测的频数(原点的指数)。所估计的分布非常不对称(在对数尺度上同样如此),这意味着正态性假设有问题。然而,在模型多估计了 9 个参数的同时,基于正态性的对数似然值( -665.29)并没有比 NPMLE 的估计值( -655.06)小很多。

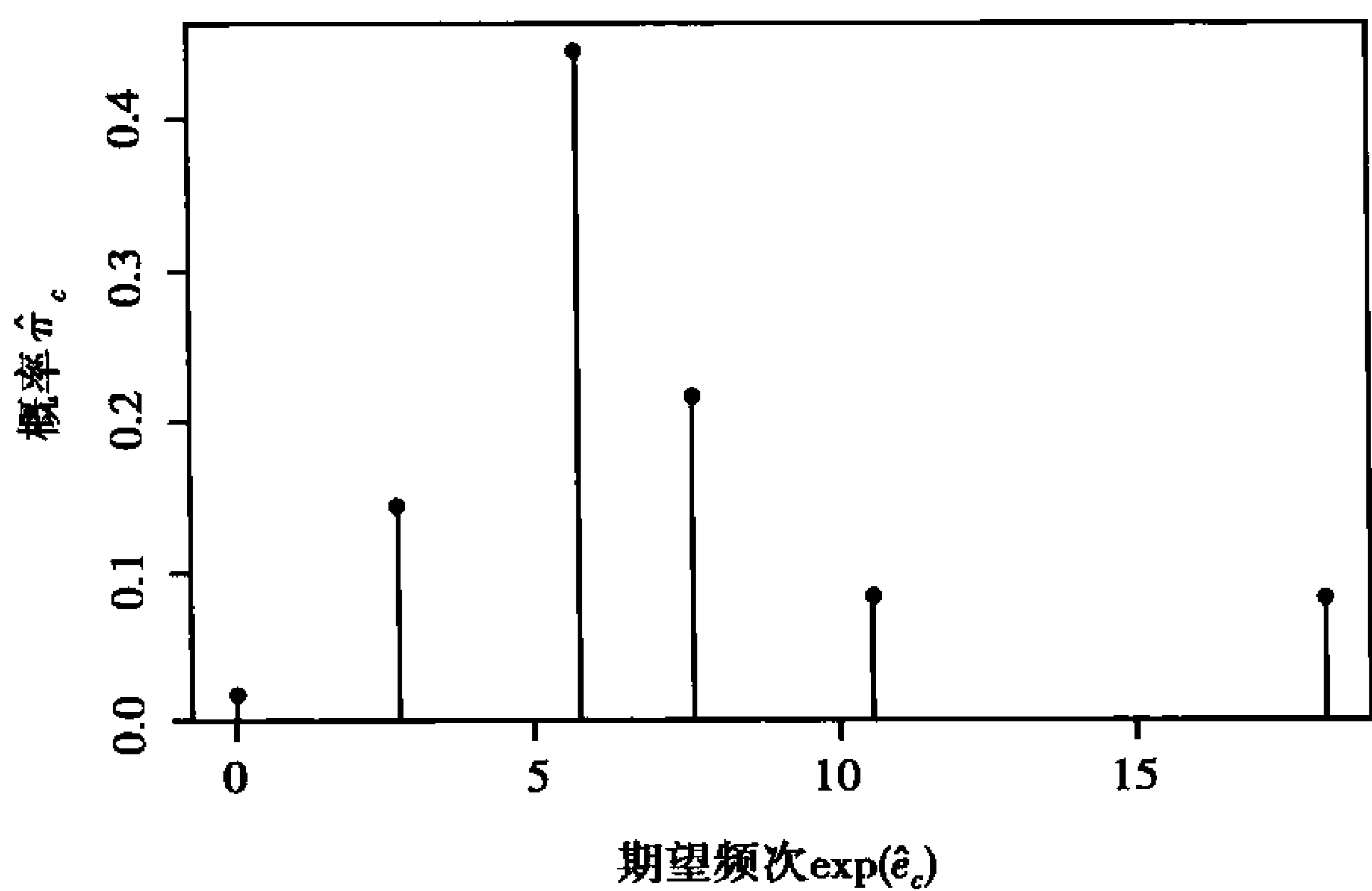



图 11.2　当 $\mathbf{x}_{ij}=\mathbf{0}$,对于 $c=1,\cdots,5$ 的概率估计 $\hat{\pi}_c$ 及期望频次 $\exp(\hat{e}_c)$

对于那两个估计频次为 0 和 18 次的质量来说,它们的存在说明可能存在着有影响力的异常值。为了评估这种影响,我们对假设为正态随机截距的模型计算了库克距离(见 8.6.6 节)。采用 7.3.2 节式(7.8)的近似抽样标准差,我们也获得了随机截距的标准化估计,即所谓的标准化经验贝叶斯残差。标准化残差通过对比当个体 $j$ 被删除后的参数估计 $\hat{\theta}_{(-j)}$ 与基于完整样本的参数估计 $\hat{\theta}$ 来计算得到。

对于库克距离大于 1 的 6 个个体,以及虽然库克距离小但是标准化残差的绝对值大于 2 的两个个体的诊断结果显示在表 11.6 中。对于[Treat]、[V4]以及随机截距标准差 $\sqrt{\psi_{11}}$,我们还给出了 DFBETAS。这是通过删除这个个体并重新估计参数得到的,而非使用 8.6.6 节的单步近似法。我们还列出了对应的响应 $y_j$,试验前的癫痫发作次数除以 4([Base]),以及试验组([Treat])。

既然一共有 59 位病人,如果残差为标准正态分布,那么标准化残差的绝对值大于 2.39的期望个数大约为 1。因而,标准化的经验贝叶斯预测大于这个值的个体就应该属于异常值(用粗体显示,如个体 225 与 232)。正如所料,个体 225 与 232 在随机截距标准差及[Treat]上都有较大的 DFBETAS([Treat]为个体间协变量)。个体 135, 227, 206 及 112 在随机截距分布上可能是异常值,并对随机截距标准差的估计有相当大影响。个体 207 有相对较小的残差且对标准差有较小的影响,尽管其响应值非常高。这可能是由于基准事件数(其对数为协变量)高于主体 225 与 232 的事件数。尽管如此,删除试验组中

的个体 207 将大大缩小所估计的试验效应,[Treat]中的高 DFBETAS 值也表明了这一点。甚至,删除个体 207 将削减交互效应[LbasTrt]。个体 126, 135 及 227 都对[V4]的估计有很大的影响。这是因为这些个体在第四次临诊的时候癫痫发作次数有显著降低。

**表 11.6 影响力统计量(库克距离与 DFBETAS)及各种残差**

| 对象 | [Base] | $y_i$ | | | | [Treat] | 库克距离 D | DFBETAS | | | 正态 | | NPML | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | [Treat] | [V4] | $\sqrt{\psi_{11}}$ | $\frac{\tilde{\zeta}_{j(-j)}}{\tilde{\sigma}_{j(-j)}}$ | $\frac{\tilde{\zeta}_j}{\tilde{\sigma}_j}$ | $\omega_1$ | $\omega_6$ |
| 126 | 13.0 | 40 | 20 | 23 | 12 | 0 | 1.10 | −0.02 | **0.52** | 0.02 | 1.04 | 0.89 | 0.00 | 0.00 |
| 135 | 2.5 | 14 | 13 | 6 | 0 | 0 | 1.52 | 0.39 | **0.40** | **−0.33** | 2.23 | 1.97 | 0.00 | **0.99** |
| 227 | 13.8 | 18 | 24 | 76 | 25 | 0 | 1.46 | −0.14 | **0.39** | **−0.33** | 2.19 | 1.93 | 0.00 | **1.00** |
| 207 | 37.8 | 102 | 65 | 72 | 63 | 1 | 1.68 | **0.58** | 0.24 | −0.16 | 1.97 | 1.37 | 0.00 | **1.00** |
| 225 | 5.5 | 1 | 23 | 19 | 8 | 1 | 1.05 | **−0.23** | 0.18 | **−0.44** | **2.47** | 2.26 | 0.00 | **1.00** |
| 232 | 3.3 | 0 | 0 | 0 | 0 | 1 | 1.57 | **0.34** | 0.00 | **−0.44** | **−2.92** | −2.77 | **0.94** | 0.00 |
| 206 | 12 | 11 | 0 | 0 | 5 | 0 | 0.52 | 0.13 | −0.08 | **−0.32** | −2.10 | −1.91 | 0.00 | 0.00 |
| 112 | 7.75 | 22 | 17 | 19 | 16 | 1 | 0.72 | −0.03 | 0.02 | **−0.32** | 2.26 | 2.07 | 0.00 | **1.00** |

对于 NPMLE 模型,我们分析最小和最大两个原点的后验概率,分别用 $\omega_1$ 与 $\omega_6$ 表示。表中除了 126 与 206 以外的其他所有个体属于其中一个极端分类的后验概率接近于 1;个体 126 与 206 属于分类 5 与分类 2 的后验概率分别为 0.99 与 1.00。至于没有显示在该表中的个体,它们在 $\omega_6$ 上最大的值为 0.03,$p_1$ 均为 0。对于各种测量"异常性"的方法,可以用模拟的方法得到 $p$ 值。

## 11.4 苏格兰唇癌:疾病制图

### 11.4.1 介 绍

接下来,我们分析疾病制图或小域估计。克雷顿与卡德(Clayton and Kaldor, 1987)展示并分析了 1975—1980 年间苏格兰 56 个县(改制前)的唇癌数据。这个数据[1]也被 Breslow and Clayton (1993)、Leyland (2001)以及其他学者分析过。观测到的唇癌病例,期望值以及粗标化死亡比(crude standardized mortality ratios (SMR))显示在表 11.7 中。

唇癌病例期望值基于苏格兰的各年龄组唇癌比率以及各县的年龄分布得出。一个县的标化死亡比(SMR)被定义为死亡率与某个特定年龄组死亡率等于参照总体死亡率时的预期死亡率的比值(如 Breslow and Day, 1987)。县 $j$ 的粗标化死亡比可通过:

$$\widehat{\text{SMR}}_j = \frac{o_j}{e_j}$$

得到,$o_j$ 为观测到的病例数,$e_j$ 为期望数。该估计值也显示在表 11.7 中,位于"粗标化死亡比"一栏下,关于粗标化死亡比的图形如图 11.3 所示。

1 该数据可从 gllamm. org/books 下载。

表 11.7 苏格兰各县唇癌病例观测值与期望值以及不同的 SMR 估计(百分比)

| 县 | # | 观测值 $o_j$ | 期望值 $e_j$ | 粗标化死亡比 | SMR 估计: 正态 | NPML | 空间 IGAR: 估计值 | 95% 置信区间 |
|---|---|---|---|---|---|---|---|---|
| Skye. Lochalsh | 1 | 9 | 1.4 | 652.2 | 470.8 | 342.6 | 412.3 | 305.5, 492.2 |
| Banf. Buchan | 2 | 39 | 8.7 | 450.3 | 421.8 | 362.4 | 430.4 | 408.4, 444.3 |
| Caithness | 3 | 11 | 3.0 | 61.8 | 309.4 | 327.1 | 351.4 | 306.7, 394.9 |
| Berwickshire | 4 | 9 | 2.5 | 355.7 | 295.2 | 321.6 | 230.3 | 162.9, 281.8 |
| Ross. Cromarty | 5 | 15 | 4.3 | 352.1 | 308.5 | 327.6 | 321.8 | 277.2, 357.3 |
| Orkney | 6 | 8 | 2.4 | 333.3 | 272.1 | 311.1 | 332.4 | 283.4, 381.9 |
| Moray | 7 | 26 | 8.1 | 320.6 | 299.9 | 322.2 | 303.2 | 275.9, 324.4 |
| Shetland | 8 | 7 | 2.3 | 304.3 | 247.8 | 292.5 | 311.1 | 274.5, 353.5 |
| Lochaber | 9 | 6 | 2.0 | 303.0 | 238.9 | 280.1 | 231.2 | 190.1, 271.0 |
| Gordon | 10 | 20 | 6.6 | 301.7 | 279.1 | 319.9 | 285.5 | 261.3, 304.4 |
| W. Isles | 11 | 13 | 4.4 | 295.5 | 262.5 | 315.5 | 299.1 | 264.7, 335.6 |
| Sutherland | 12 | 5 | 1.8 | 279.3 | 219.2 | 254.3 | 304.1 | 261.7, 354.9 |
| Nairn | 13 | 3 | 1.1 | 277.8 | 198.4 | 222.7 | 266.2 | 215.0, 317.0 |
| Wigtown | 14 | 8 | 3.3 | 241.7 | 210.9 | 249.6 | 159.7 | 115.0, 194.6 |
| NE. Fife | 15 | 17 | 7.8 | 216.8 | 204.6 | 245.3 | 173.2 | 140.5, 194.0 |
| Kincardine | 16 | 9 | 4.6 | 197.8 | 178.9 | 171.4 | 190.9 | 167.9, 214.5 |
| Badenoch | 17 | 2 | 1.1 | 186.9 | 151.9 | 163.2 | 207.7 | 169.7, 247.8 |
| Ettrick | 18 | 7 | 4.2 | 167.5 | 154.7 | 136.7 | 126.9 | 102.9, 153.5 |
| Inverness | 19 | 9 | 5.5 | 162.7 | 154.2 | 128.4 | 210.4 | 179.7, 252.3 |
| Roxburgh | 20 | 7 | 4.4 | 157.7 | 149.0 | 130.3 | 145.0 | 120.2, 172.6 |
| Angus | 21 | 16 | 10.5 | 153 | 147.8 | 117.1 | 138.8 | 125.2, 150.1 |
| Aberdeen | 22 | 31 | 22.7 | 136.7 | 135.0 | 116.4 | 145.0 | 139.0, 156.8 |
| Argyll. Bute | 23 | 11 | 8.8 | 125.4 | 123.3 | 116.4 | 111.5 | 96.5, 126.1 |
| Clydesdale | 24 | 7 | 5.6 | 124.6 | 122.9 | 116.7 | 79.7 | 63.7, 95.0 |
| Kirkcaldy | 25 | 19 | 15.5 | 122.8 | 121.6 | 116.4 | 123.9 | 115.1, 132.1 |
| Dunfermline | 26 | 15 | 12.5 | 120.1 | 119.1 | 116.3 | 108.4 | 94.6, 118.8 |
| Nithsdale | 27 | 7 | 6.0 | 115.9 | 116.3 | 115.5 | 99.8 | 79.9, 115.2 |
| E. Lothian | 28 | 10 | 9.0 | 111.6 | 111.4 | 115.9 | 108.7 | 94.7, 123.6 |
| Perth. Kinross | 29 | 16 | 14.4 | 111.3 | 111.1 | 116.3 | 116.9 | 105.4, 129.5 |
| W. Lothian | 30 | 11 | 10.2 | 107.8 | 108.5 | 115.9 | 78.9 | 64.9, 91.1 |
| Cumnock-Doon | 31 | 5 | 4.8 | 105.3 | 107.2 | 111.5 | 91.2 | 74.0, 108.4 |
| Stewartry | 32 | 3 | 2.9 | 104.2 | 109.1 | 109.4 | 107.4 | 82.6, 131.7 |

续表

| | | | | | SMR 估计 | | | |
|---|---|---|---|---|---|---|---|---|
| | | 观测值 | 期望值 | 粗标化 | | | 空间 IGAR | |
| 县 | # | $o_j$ | $e_j$ | 死亡比 | 正态 | NPML | 估计值 | 95% 置信区间 |
| Midlothian | 33 | 7 | 7.0 | 99.6 | 102.7 | 113.1 | 90.4 | 75.1, 104.3 |
| Stirling | 34 | 8 | 8.5 | 93.8 | 97.3 | 112.9 | 74.2 | 60.8, 85.7 |
| Kyle. Carrick | 35 | 11 | 12.3 | 89.3 | 92.2 | 114.1 | 91.1 | 79.1, 104.5 |
| Inverclyde | 36 | 9 | 10.1 | 89.1 | 92.5 | 112.4 | 84.4 | 73.7, 95.7 |
| Cunninghame | 37 | 11 | 12.7 | 86.8 | 89.7 | 113.3 | 81.9 | 71.4, 90.9 |
| Monklands | 38 | 8 | 9.4 | 85.6 | 89.5 | 109.4 | 63.2 | 50.1, 74.1 |
| Dumbarton | 39 | 6 | 7.2 | 83.3 | 89.4 | 104.9 | 78.1 | 66.0, 90.8 |
| Clydebank | 40 | 4 | 5.3 | 75.9 | 85.6 | 94.5 | 61.5 | 50.4, 73.0 |
| Renfrew | 41 | 10 | 18.8 | 53.3 | 59.1 | 40.6 | 56.6 | 49.6, 64.8 |
| Falkirk | 42 | 8 | 15.8 | 50.7 | 57.9 | 40.9 | 61.8 | 53.8, 70.7 |
| Clackmannan | 43 | 2 | 4.3 | 46.3 | 68.8 | 65.5 | 75.1 | 63.0, 89.8 |
| Motherwell | 44 | 6 | 14.6 | 41.0 | 50.6 | 37.5 | 50.1 | 43.5, 57.6 |
| Edinburgh | 45 | 19 | 50.7 | 37.5 | 40.8 | 36.2 | 46.6 | 41.8, 55.9 |
| Kilmarnock | 46 | 3 | 8.2 | 36.6 | 53.2 | 42.2 | 56.7 | 47.5, 67.6 |
| E. Kilbride | 47 | 2 | 5.6 | 35.8 | 57.9 | 49.7 | 51.2 | 40.6, 61.5 |
| Hamilton | 48 | 3 | 9.3 | 32.1 | 48.5 | 38.8 | 45.4 | 38.3, 54.8 |
| Glasgow | 49 | 28 | 88.7 | 31.6 | 33.8 | 36.2 | 37.0 | 33.8, 41.2 |
| Dundee | 50 | 6 | 19.6 | 30.6 | 39.8 | 36.2 | 57.4 | 46.3, 77.9 |
| Cumbernauld | 51 | 1 | 3.4 | 29.1 | 63.1 | 57.8 | 54.0 | 43.4, 66.8 |
| Bearsden | 52 | 1 | 3.6 | 27.6 | 61.1 | 55.4 | 48.3 | 37.9, 60.3 |
| Eastwood | 53 | 1 | 5.7 | 17.4 | 46.4 | 40.6 | 42.2 | 34.5, 52.7 |
| Strathkelvin | 54 | 1 | 7.0 | 14.2 | 40.8 | 37.8 | 43.2 | 33.9, 53.6 |
| Tweeddale | 55 | 0 | 4.2 | 0.0 | 43.2 | 40.8 | 73.0 | 55.1, 97.0 |
| Annandale | 56 | 0 | 1.8 | 0.0 | 64.9 | 60.6 | 71.3 | 56.0, 87.1 |

资料来源:Clayton and Kaldor (1987)。

关于粗标化死亡比存在两个非常重要的限制。首先,用粗标化死亡比对人口数很少的县进行估计,其结果会非常不准确。其次,粗标化死亡比并不考虑地理上接近的区域会具有相似的发病率。第一个问题可以通过随机截距泊松模型,得到缩小的标化死亡比贝叶斯或经验贝叶斯估计,从而从其他县"借力"。第二个问题可以通过允许随机截距在空间上相关来解决。

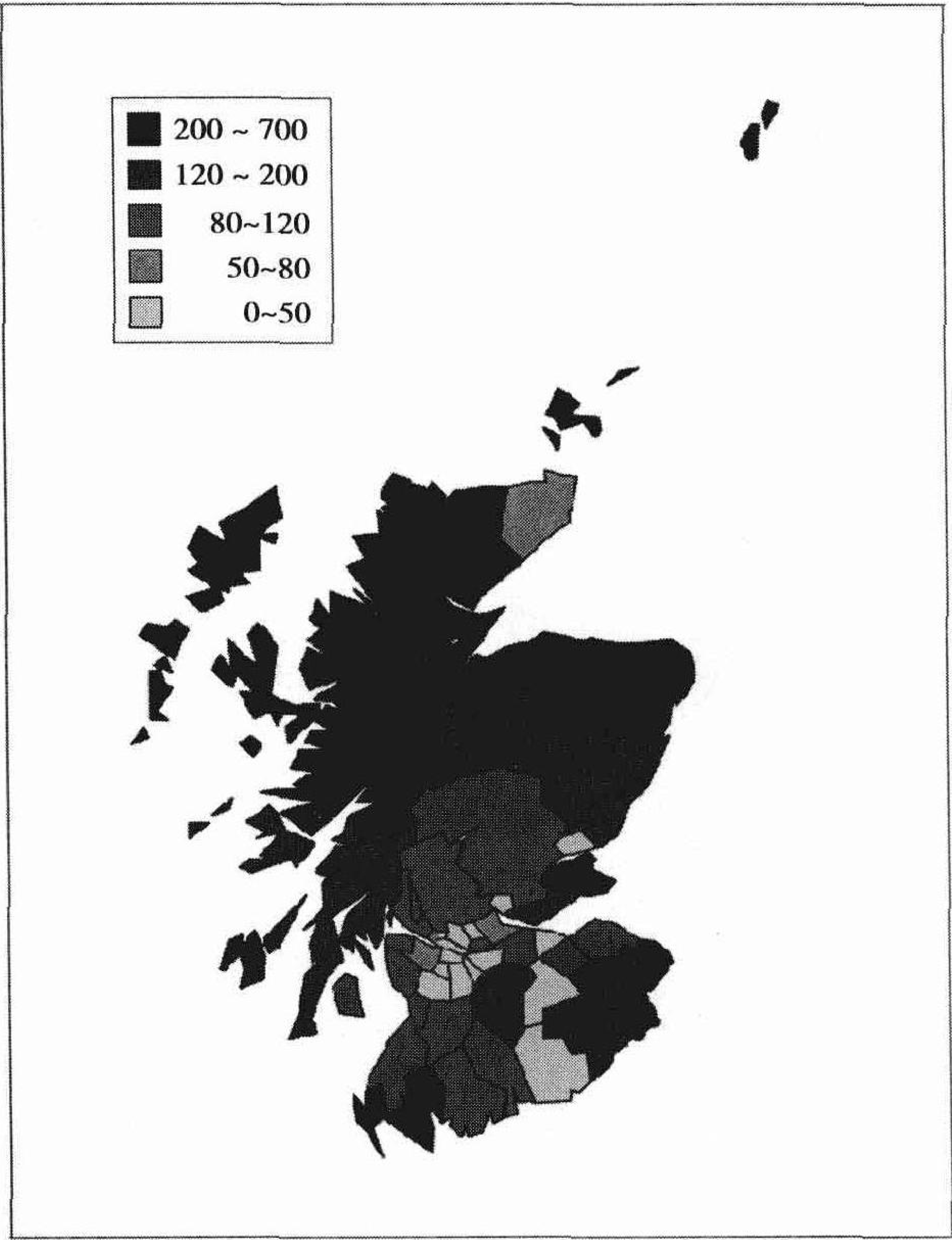



图 11.3 粗标化死亡比(百分比)

## 11.4.2 随机截距模型

我们首先讨论一个简单随机截距模型。假设在 $j$ 县所观测到的唇癌病例符合均值为 $\mu_j$ 的泊松分布,即

$$\ln(\mu_j) = \ln(e_j) + \beta_0 + \zeta_j$$

其中,$\ln(\mu_j)$ 为偏移量(回归系数设为 1 的系数),$\zeta_j$ 为随机截距,表示各县之间未观测到的异质性。标化死亡比的经验贝叶斯估计

$$\widetilde{\mathrm{SMR}}_j = \mathrm{E}_\zeta[\exp(\hat{\beta}_0 + \zeta_j) \mid o_j, e_j]$$

将缩减为标化死亡比的均值,并对人口数较少的县也能得到稳定的估计值。

在 $\zeta_j \sim \mathrm{N}(0,\psi)$ 假设下的参数估计显示在表 11.8 中的"独立正态"栏下。在 $\zeta_j \sim \mathrm{N}(0,\psi)$ 假设下的标化死亡比的经验贝叶斯估计显示在表 11.7 中的"正态"栏下,其图形如图 11.4 所示。

除了假设随机截距正态分布之外,我们也可以使用非参数最大似然估计(NPMLE),其中,$\zeta_j$ 是原点为 $\zeta_j = e_c, c = 1, \cdots, C$,概率为 $\pi_c$ 的离散值,$C$ 将使似然值最大化(见 6.5 节)。NPML 解有 $C=4$ 个质量,其原点 $\hat{\beta}_0 + \hat{e}_c$ 分别等于 -1.02, 0.15, 1.13 及 1.35(对应于当 SMR 等于 36.2%, 116.4%, 308.8% 及 384.8% 时),估计的概率 $\hat{\pi}_c$ 等于27.5%, 47.9%, 18.5% 及6.2%。参数估计显示在表 11.8 的"独立 NPML"栏中。根据这几个 NPML 估计值所预测的标化死亡比显示在表 11.7 的"NPML"栏中。

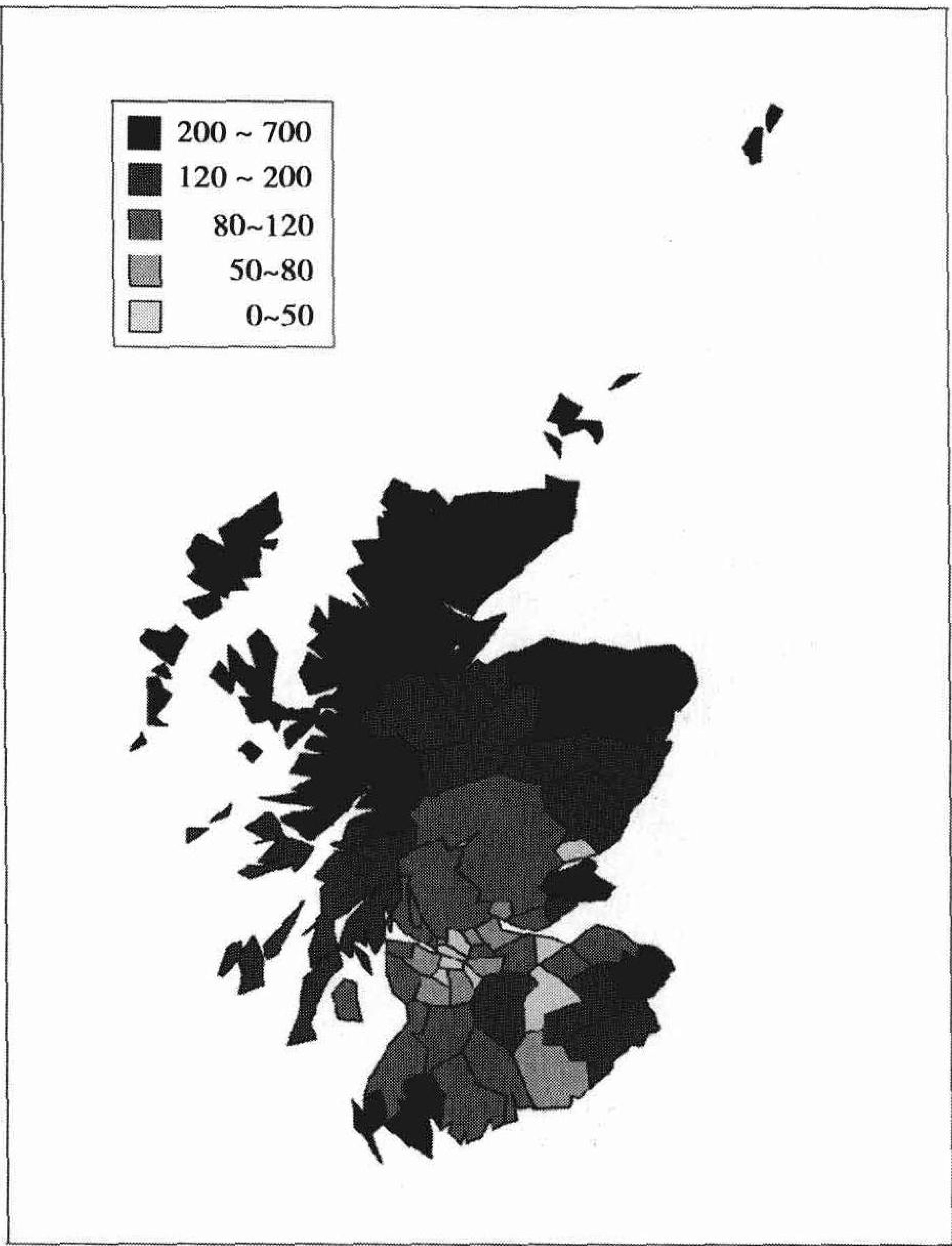



**图 11.4 假设为正态分布随机效应的标化死亡比(无协变量)**

**表 11.8 苏格兰唇癌数据的几种随机截距模型参数估计**

| | 独立正态 | | 独立 NPML | | 空间 IGAR | |
|---|---|---|---|---|---|---|
| | Est | (SE) | Est | (SE) | Est | (SE) |
| $\beta_0$[Cons] | 0.08 | (0.12) | 0.08 | (0.12) | 0.09 | (0.07) |
| $\sqrt{\psi}$ | 0.76 | (0.10) | $0.80^{\dagger}$ | (—) | 0.74 | (0.14) |

注:†来自离散分布。

## 11.4.3 空间相关的随机截距

未观测到的风险因素,尤其是环境风险因素,在空间上很可能是相关的。因此,我们现在来分析空间模型,即允许随机效应 $\zeta_j$ 与相邻的几个县相关。克雷顿与卡德(Clayton and Kaldor, 1987)讨论了条件高斯自回归模型(conditional Gaussian autoregressive model)(Besag, 1974),但是,在本例子中因相邻的个数不同,也即被称为“不规则图形”,故使用这种模型设定将会有问题。因而,对于这种情形,内在自回归高斯模型(intrinsic autoregressive Gaussian, IGAR)被认为更适合(Besag et al., 1991; Bernardinelli and Montomoli, 1992)。在这个模型里,在给定相邻县的随机截距条件下,某个县的随机截距的条件分布并不依赖于非相邻县的随机截距。这种模型也被称为马尔可夫随机场。该模型被设定为

$$h(\zeta_j) \propto \exp \sum_{i \sim j} (\zeta_j - \zeta_i)^2/(2\psi)$$

其中,$i \sim j$ 表示与 $j$ 县相邻的县。在此,$\zeta_j$ 的条件期望值为相邻各县的 $\zeta_i$ 的均值,即

$$E(\zeta_j \mid \zeta_i, i \neq j) = \frac{1}{m_j} \sum_{i \sim j} \zeta_i$$

请注意,并未设定 $\zeta_j$ 的非条件期望值。条件方差与其个数 $m_j$ 成反比,即

$$\mathrm{Var}(\zeta_j \mid \zeta_i, i \neq j) = \frac{\psi}{m_j}$$

该唇癌数据的内在自回归高斯模型的估计采用 PQL-1 (Breslow and Clayton, 1993)与 MCMC (Spiegelhalter et al., 1996a)估计法。由于所有县的随机效应两两相关(除了海岛县),基于数值积分的方法就不再适用于这种高积分维度的情形。因而我们提出下述非常类似于6.11.5 节所讨论的 AIP 算法的迭代算法:

- 设置初始偏移量为零,则

$$a_j^0 = 0$$

- 在迭代 $k$ 中:

(1)给定参数估计 $\hat{\boldsymbol{\theta}}^k$ 及其协方差矩阵估计 $\mathrm{Cov}(\hat{\boldsymbol{\theta}})^k$,估计一个带线性估计量的随机截距泊松模型:

$$\nu_j = \ln(e_j) + \beta_0 + a_j^k + u_j \frac{1}{\sqrt{m_j}}, \quad u_j \sim \mathrm{N}(0, \psi)$$

(2)根据近似样本分布对参数 $\boldsymbol{\theta}^k$ 抽样,则

$$\boldsymbol{\theta}^k \sim \mathrm{N}(\hat{\boldsymbol{\theta}}^k, \mathrm{Cov}(\hat{\boldsymbol{\theta}})^k)$$

(3)给定参数 $\boldsymbol{\theta}^k$,得到随机效应的后验均值与标准偏差,即

$$\hat{u}_j^k = E(u_j \mid \mathbf{y}_j, \boldsymbol{\theta}^k)$$
$$\tau_j^k = \mathrm{Var}(u_j \mid \mathbf{y}_j, \boldsymbol{\theta}^k)$$

(4)从近似后验分布中对随机效应抽样,则

$$u_j^k \sim \mathrm{N}(\tilde{u}_j^k, \tau_j^k)$$

(5)得到 $\zeta_j^k$

$$\zeta_j^k = a_j^k + u_j^k \frac{1}{\sqrt{m_j}}$$

(6)计算相邻县 $\zeta_j^k$ 的均值为

$$b_j^k = \frac{1}{m_j} \sum_{i \sim j} \zeta_i^k$$

(7)偏移量即为

$$a_j^{k+1} = b_j^k - \frac{1}{J} \sum_{j=1}^{J} b_j^k$$

在此,偏移量 $a_j^k$ 表示在当前迭代中第 $j$ 个组的随机效应 $\zeta_j^k$ 的均值,是对前一次迭代中相邻县的随机效应的均值计算而得(为了辨识截距参数 $\beta_0$,在第七步中的均值中心化是必需的,因为内在自回归高斯模型并不设定随机效应的均值)。在第一步中,$\zeta_j = a_j^k + u_j \frac{1}{\sqrt{m_j}}$ 均值为 $a_j^k$,方差为 $\psi/m_j$。随机效应 $u_j$ 为独立分布,因而模型可用任何估计随机截距模型的软件来进行估计,在此采用 gllamm 程序的自适应积分法。在第二步中,为了反映估计的不确定性,我们根据近似样本分布进行抽样。在传统贝叶斯法中参数估计是替代进来的,且不确定性被忽略掉了,因而,我们的方法比传统的经验贝叶斯法更加精细。在第三步与第四步中,我们根据近似正态分布对随机效应进行抽样得到后验分布。在第

五步中,我们得到对应的 $\zeta_j^k$,并在第六步中将其合并得到下一步必需的偏移量。在第七步中,对偏移量进行了均值中心化以对截距 $\beta_0$ 进行估计。

计算迭代一直进行,直到估计值 $\hat{\theta}^k$ 看上去来自一个比较稳定的分布(迭代期)。本例仅进行了大约10次迭代。我们用该算法重复进行了500次计算,以得到均值及估计值的方差,正如6.11.5节所讨论的。估计结果如表11.8所示。

在第3步中,得到了标化死亡比的后验均值。这些均值及基于500次迭代的2.5与97.5百分点的95%置信区间显示在表11.7的最后3列中。点估计值显示在图11.5的地图中,其置信区间显示在图11.6中。空心圆圈为粗标化死亡比估计值,垂直线为使用内在高斯自回归模型所估计的标化死亡比均值,$100\exp(\hat{\beta}_0)$。

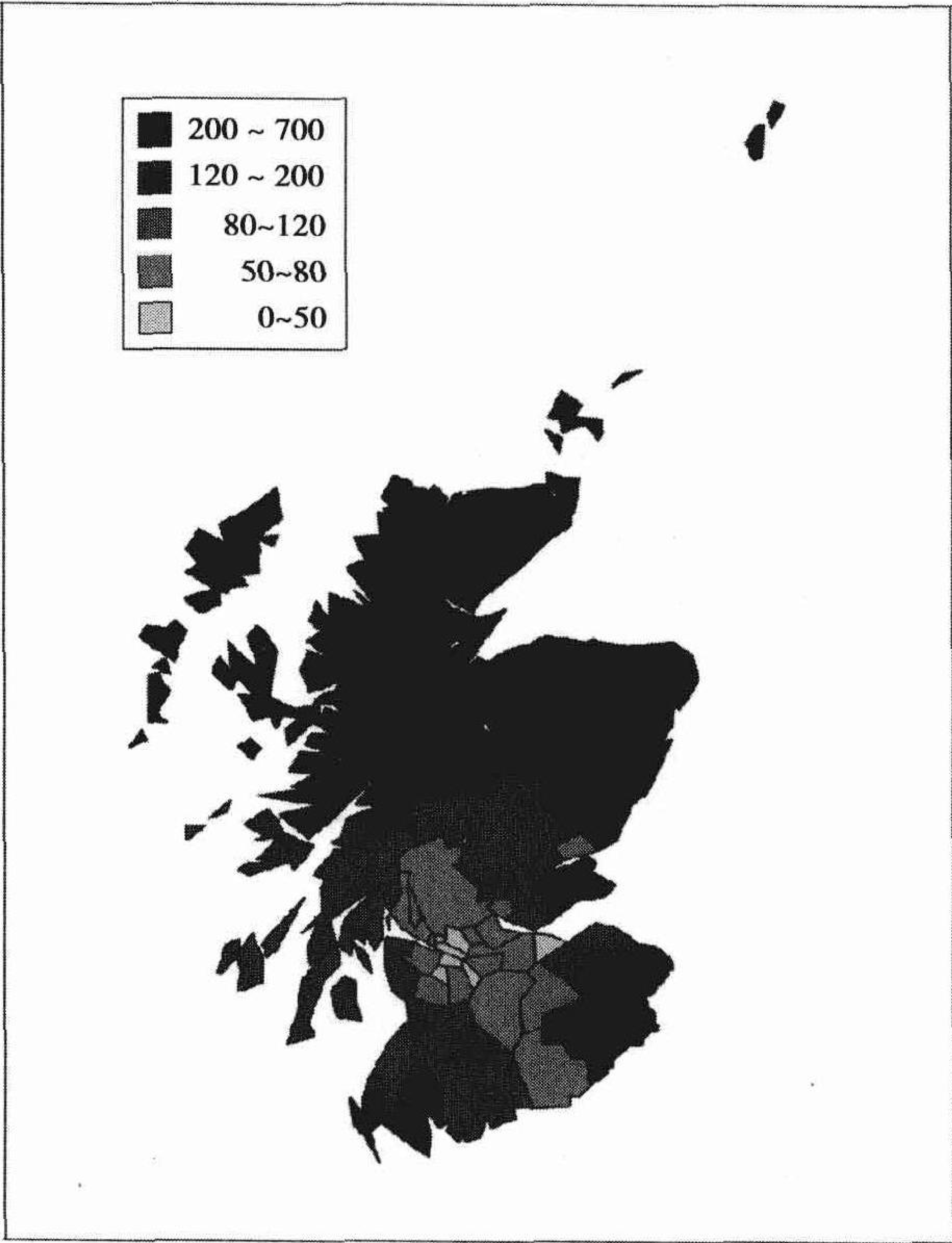



**图11.5 各县的标化死亡比及基于不含协变量的内在自回归高斯模型的95%置信区间估计**

## 11.4.4 引入县层次的协变量

布里斯罗和克雷顿(Breslow and Clayton, 1993)提出分析县层次的协变量[Agric](每个县中,农业、林业及渔业劳动力所占比例除以10)的效应。由于这3种职业都需要进行户外工作,而阳光被认为是唇癌的主要风险因素(Kemp et al., 1985),因而[Agric]对唇癌应该存在效应。根据布里斯罗和克雷顿(Breslow and Clayton, 1993),我们分析下述模型:

$$\ln(\mu_j) = \ln(e_j) + \beta_0 + \beta_1 x_j + \zeta_j$$

其中,[Agric]用 $x_j$ 表示。

包含[Agric]效应的内在高斯自回归模型的估计采用数据增强算法(data augmentation algorithm),估计结果显示在表11.9中。这个估计结果与采用PQL-1

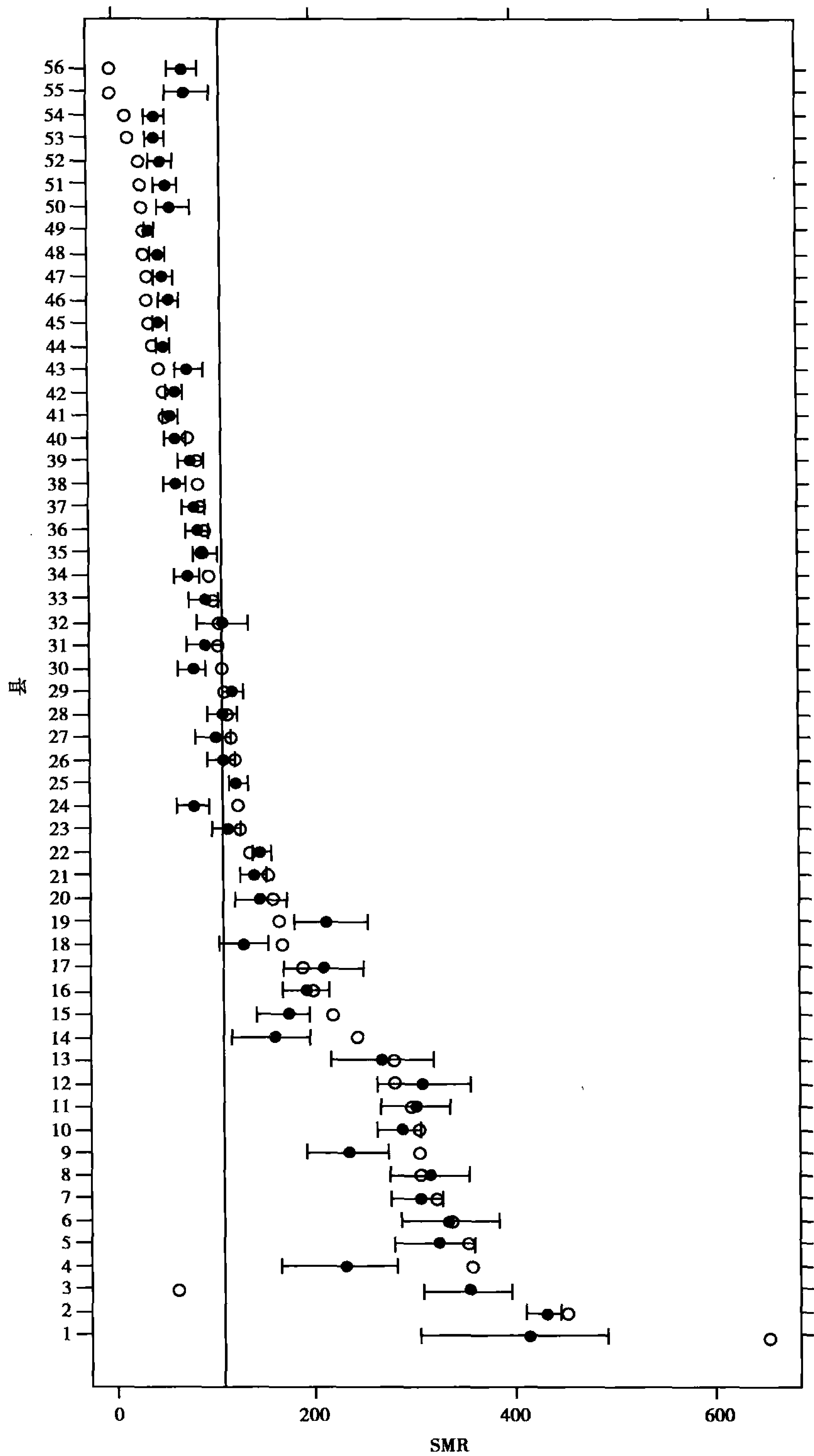



**图 11.6 不含协变量的内在高斯自回归模型估计的标化死亡比近似 95% 置信区间**

(Breslow and Clayton, 1993)及 MCMC 法（根据 BUGS 手册例子(Spiegelhalter et al., 1996c,第 2 卷, 第 11 章)的估计结果非常接近,3 种方法的估计结果均显示在表格中。贝叶斯模型的先验为 $\beta_1 \sim N(0,10^5)$ 及 $\zeta_j \sim N(0,\psi)$,超先验为 $\psi \sim IG(0.001,0.001)$,请

见6.11.5节的**逆伽玛分布**(IG)。MCMC估计法并没有包含常数项,因为斯皮格哈特等(Spiegelhalter et al., 1996c)并未将随机效应的均值设为零。尽管看上去存在[Agric]的效应,但是当基于像本例子这样的生态学或聚合数据来对病源下结论时仍需高度谨慎。

**表11.9 包含协变量的内在高斯自回归模型对苏格兰唇癌数据的估计**

| | PQL-1 | | MCMC | | 数据增强 | |
|---|---|---|---|---|---|---|
| | Est | (SE) | Est | (SE) | Est | (SE) |
| $\beta_0$[Cons] | -0.18 | (0.12) | — | | -0.21 | (0.11) |
| $\beta_1$[Agric] | 0.35 | (0.12) | 0.37 | (0.11) | 0.36 | (0.11) |
| $\sqrt{\psi}$ | 0.73 | (0.13) | 0.69 | (0.12) | 0.65 | (0.14) |

在图11.7中,显示的是基于包含协变量[Agric]的内在高斯自回归模型采用数据增强算法估计的标化死亡比地图。

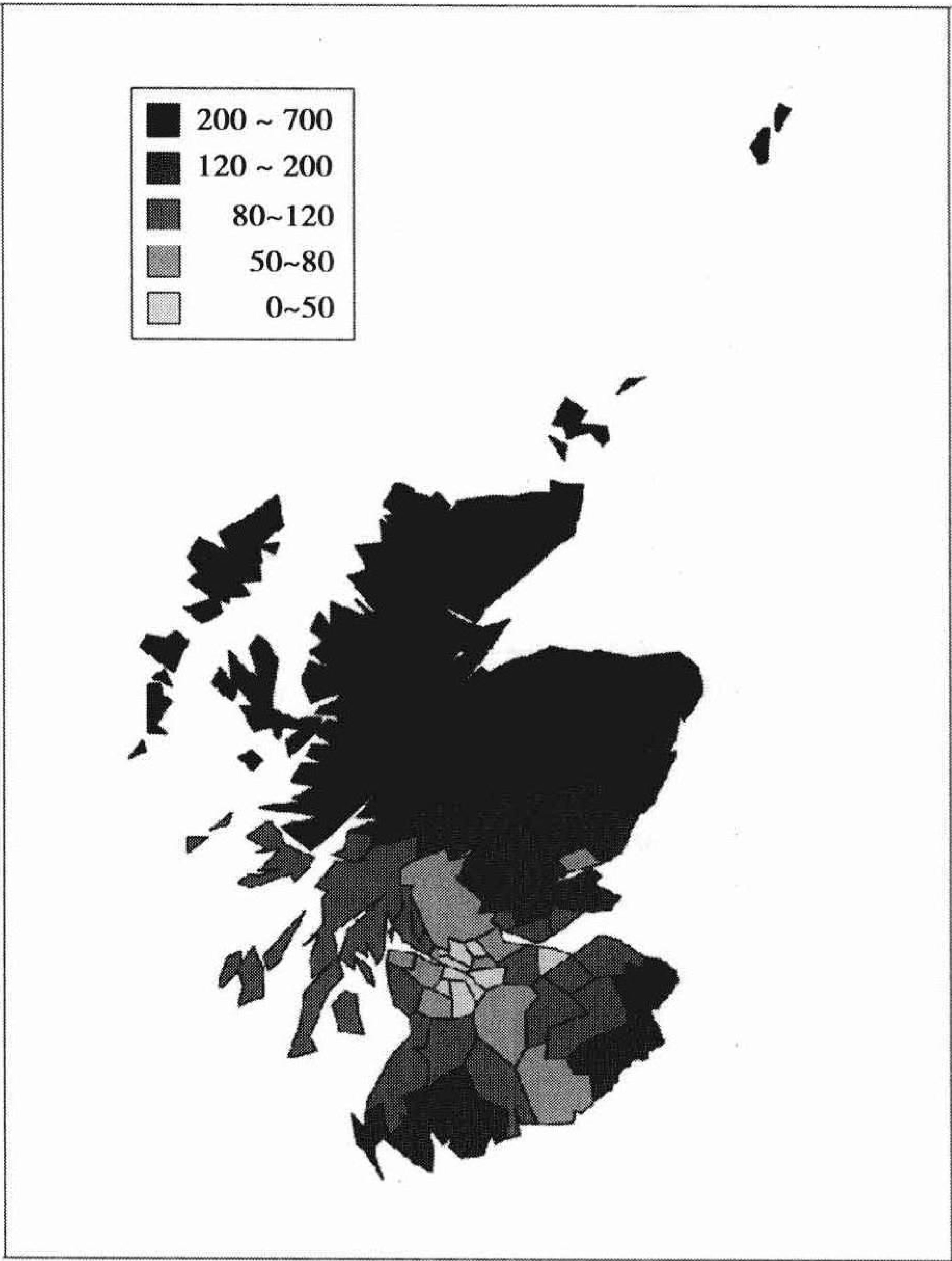



**图11.7 包含协变量[Agric]的内在自回归高斯模型估计的标化死亡比**

## 11.5 总结与进一步阅读

我们讨论了过度离散的问题,采用零堆积模型以及包含正态分布和非参随机效应分布的随机截距模型。尽管这些模型来源于统计学文献(Lambert, 1992),但近来关于零堆积数据模型的研究及其应用主要来自计量经济学(请参见Cameron and Trivedi (1998))。Hall (1997)及Dobbie and Welsh (2001)针对群组数据讨论了零堆积泊松模型。

接着,我们使用随机截距模型以及随机系数模型分析了纵贯性癫痫症数据,并对模

型假设进行判定。对于各种类型的事件数数据模型及包含随机效应的模型的处理可参见 Winkelmann (2003)及 Cameron and Trivedi (1998)的计量经济学专著。

对于唇癌数据,我们使用了不同的模型(包括空间相关随机效应模型)来进行制图分析。对于空间依赖的情形,还可采用条件依赖于相邻事件数的条件自回归模型。比杰里等(Biggeri et al., 2000)将随机效应包含到模型中,并采用非参最大似然估计法。诺赫德和贝斯特(Knorr-Held and Best, 2001)针对两种疾病提出了一个贝叶斯模型,在这个模型中,响应模型实际上是公共因子模型。小域估计以及制图的应用主要在医学研究中,如将发病率显示在地图上。然而,其他的应用也可将诸如失业率、犯罪率等显示在地图上。普费弗曼(Pfeffermann, 2002)提供了一个很好的综述。罗森等 (Lawson et al., 2003)讨论用软件 MLwiN 及 WinBUGS 来进行疾病制图。

在本章中,我们没有使用因子模型。关于事件数数据的因子模型的讨论及应用请参见维德尔等 (Wedel et al., 2003)。阿明戈尔和库斯特(Arminger and Küsters, 1988)讨论了针对包含事件数数据的混合响应变量的因子模型。

在 14.5 节讨论多过程及混合响应变量时,我们将再次讨论事件数变量。我们将在考虑未观测到的异质性及内生性因素的同时,分析医生建议对饮酒量的模型效应。

# 12 存续与生存响应

## 12.1 介　绍

2.5 节考察了存续响应的模型,区分了连续时间存续与离散时间存续。在这一章中,将对群集的生存或者存续数据进行讨论,即常说的“多元”存续数据。它对区分“单一事件”和“多重事件”群集存续数据有较大帮助。

**单一事件群集存续数据**(Single events clustered duration data)由不同群集中(如家庭)每个单元(如受试者)某个事件(如因肺癌致死)的存续组成。通常假设这类事件是“吸收性的”(仅能发生一次),另外,某个单元的风险独立于在同一个群集内的其他单元的事件,但一个群集内单元之间的风险是相关的。

相反,**多重事件群集存续数据**(multiple events clustered duration data)由每个群集(如受试者)的多个事件的存续所组成。在这种情况下,某个受试者的一个事件的风险依赖于此受试者其他事件发生的时间。如果事件不同,并且是吸收性的(如由各种类型癌症导致的死亡),那么事件就是**多重类型**的。如果同样的事件(如结肠癌症发生,随后再次发生)能被重复观测到,那么则称多重事件为“**循环发生的**(recurrent)”。复杂的存续数据理所当然包括多重类型的循环事件。

在 12.2 节,对多重事件群集存续数据的模型设定进行了讨论。在 12.3 节,我们应用了离散时间单一事件的群集存续数据模型。数据是关于以学校为群集的在校学生,所关心的事件是初次接触香烟的时间。在 12.4 节,我们将模型应用到连续时间的多重事件群集生存数据模型。将分析一个随机对照试验,其响应是在重复的锻炼测试中,心绞痛发作的次数。最后,除了离散潜变量以外,还将考察连续随机效应和因子。

## 12.2 多重事件群集存续数据的建模

在 2.5 节论述的存续模型中,单一事件群集存续数据建模直接要涉及诸如随机效应和因子(可能在多个层次上变动)在内的潜变量的问题。然而,对于多重事件群集生存数据来说,在选择适当的模型时,还需要考虑以下这些问题:

(1)这些事件是否属于不同类型?

(2)这些事件是否有顺序,使得只有前一个事件已经发生,该个体才具有发生下一个事件的风险?比如 HIV 感染与 AIDS。而疟疾的例子则显得不那么清楚,因为导致第一次感染的生物过程与随后的再次发生是非常不同的。

(3)一个事件的风险是否取决于另一个事件已经在同一个个体上发生(“状态依赖”)?

(4)所有事件的风险是从一个共同的时间起点平行发展,还是从前一个事件顺序发展的? 例如,不同类型传染病的风险可能是从一个共同的起点,如手术时间,不断变化发展而来的。相反,对于重复试验的存续,如自开始锻炼(休息之后)到心绞痛发展的时间,风险可能从每个试验的开始就已经开始(见12.4节中的应用例子)。

在2.5节,连续时间存续的比例风险模型和离散时间风险模型(discrete-time hazards models)都可表示为风险集合。循环事件数据的一个难点在于,有许多不同的建构风险集合的方法,究竟哪个方法更为合适取决于上面所列举的考虑问题。

凯利和里姆(Kelly and Lim,2000)认为,在建构风险集合时下面几个方面很有用处:

• 风险间隔表示法

——总时间:时钟从观测开始起计算,不受事件发生所干扰。

——过程计数:时钟继续运行,但是一个个体只有在经历了第 $k-1$ 个事件后,才可能处于第 $k$ 个事件发生的风险中。

——间隙时间:对一个单元来说,每次事件时间之后都将时钟设置为0。

用总时间、过程计数和间隙时间来表示风险间隔,这其中的差异我们可用一个例子清晰地展现出来。考察3个主体 $j=A$, B, C,其可能的循环存续为 $t_{ij}$,即从观测开始到第 $i$ 个事件发生($\delta_{ij}=1$)或者删失($\delta_{ij}=0$)。数据显示在表12.1中,其中我们可以看到,A在时间2经历了第一个事件,在时间5经历了第二个事件,最终在时间14删失。用这3种表示法所表示的风险间隔显示在图12.1中。

**表12.1 循环存续数据举例**

| $j$ | $i$ | $t_{ij}$ | $\delta_{ij}$ |
|---|---|---|---|
| A | 1 | 2 | 1 |
| A | 2 | 5 | 1 |
| A | 3 | 14 | 0 |
| B | 1 | 7 | 1 |
| B | 2 | 12 | 1 |
| B | 3 | 17 | 0 |
| C | 1 | 14 | 1 |

风险间隔表示法的选择要根据上面问题列表中的第4个问题。如果风险来自于共同的起点,那么通常会采用总时间或者过程计数表示法;如果风险发展是顺序而来的,那么通常会采用间隔时间表示法。

• 基准风险(Baseline hazard)(单元内分化)

——公共基准风险:假定所有循环事件具有相同的基准风险 $h^0(t_{ij})$。

——特定事件基准风险:对于每个事件 $i$ 设定不同的基准风险 $h_i^0(t_{ij})$。

基准风险的选择要根据上面问题列表中的第一个问题。如果事件都是同一类型,通常设定公共基准风险;如果事件为不同的类型,通常设定特定事件基准风险。

• 第 $k$ 个事件的风险集合(单元间分化)

——受限的:对第 $k$ 个风险集合的贡献被限定为经历过 $k-1$ 个事件的那些单元。

——未受限的:风险集合包括所有仍处于风险的单元,无论它们之前已经经历了多少个事件。

——准受限的:对第 $k$ 个风险集合的贡献被限定为经历了少于 $k$ 个事件的单元。

这个选择要根据第二个问题(即是否只有当前一个事件已经发生,该个体才处在下

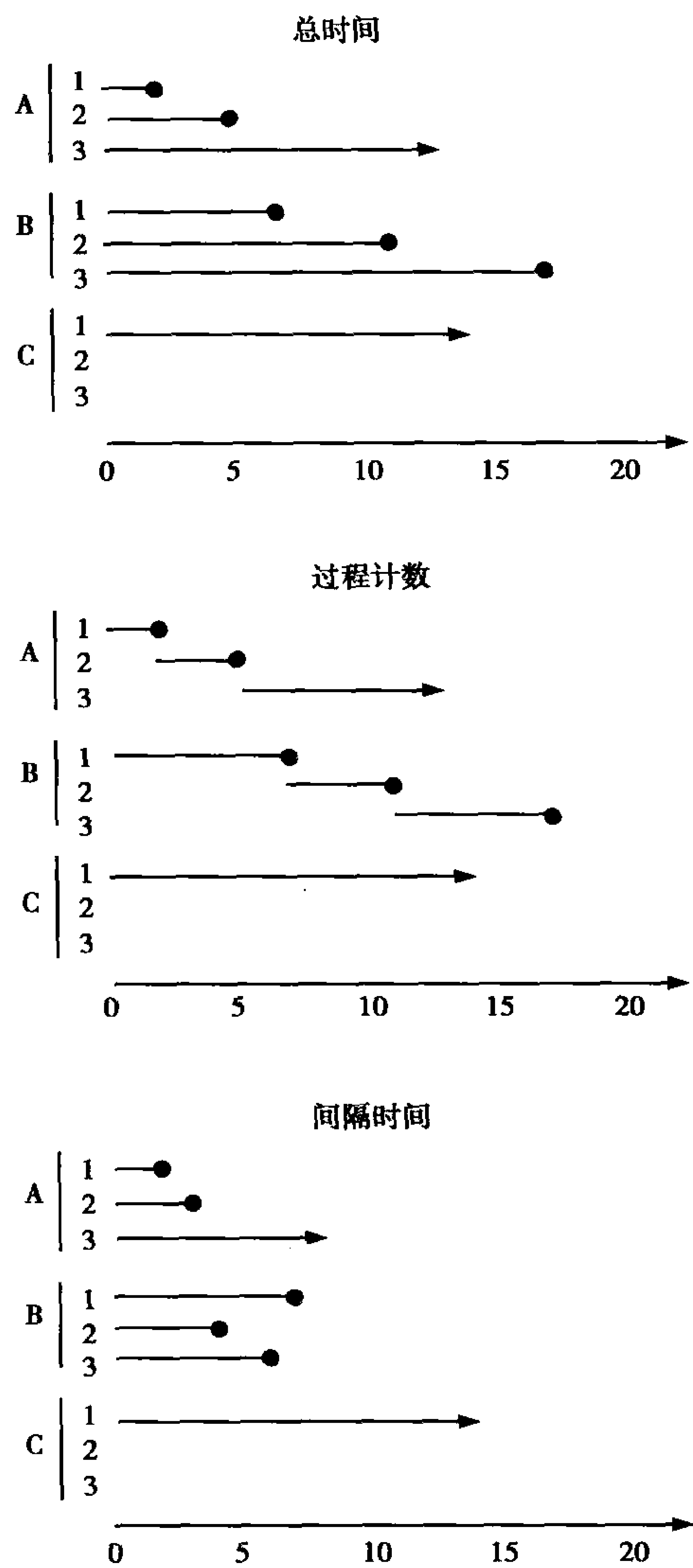



**图 12.1 用总时间、过程计数和间隔时间表示的风险间隔(引自 Kelly and Lim,2000)**

一个事件发生的风险中)。

对于连续时间比例风险建模,表 12.2 给出了各种风险间隔、风险集合和基准风险的选择组合及其关键文献。

**表 12.2 不同类型的风险间隔、风险集合和基准风险及其参考文献**

| | 风险集合/基准风险 | | |
|---|---|---|---|
| 风险间隔 | 未受限的/公共的 | 准受限的/事件-特定 | 受限的/事件-特定 |
| 总时间 | Lee, Wei & Amato (1992) | Wei, Lin & Weissfeld (1989) | |
| 过程计数 | Andersen & Gill(1982) | | Prentice, Williams & Peterson(1981) |
| 间隔时间 | | Not possible | Prentice, Williams & Peterson(1981) |

资料来源:凯利和里姆(Kelly and Lim, 2000)。

## 12.3 吸烟的肇端：离散时间脆弱模型

### 12.3.1 介 绍

这里要分析关于校园中青少年吸烟行为的数据(Lader and Matheson, 1991),这个数据之前由皮科尔斯等(Pickles et al, 2001)和拉贝-赫斯基思等(Rabe-Hesketh, 2001d)分析过。1990—1993年,在学校开展了对年龄在11~15岁的学生的两次横截面调查,得到了一个时间序列的设计(请参见 Appelbaum and McCall,1983),两个研究都遵循了相似的二阶抽样设计,将学校作为初级抽样单元。1990年的样本包括来自于125所学校的3 124名学生,1993年的样本包括来自110所学校的3 140名学生。假设学校的抽样比率足够低,这样可以忽略同一所学校两次出现在调查中的情况。由于在每次调查中都抽取了5个年级的班级,因此两次调查间隔3年可能会导致一些年龄群被抽中两次(比如,在1990年11岁的人与在1993年14岁的人属于同一个年龄群)。

学生们被问到是否抽过烟,如果抽过,则记录其第一次吸烟时的年龄。在观测数据中,2%的学生不记得他们第一次吸烟时的年龄(左侧删失),这些数据被我们去除了。

对这些学生的访问是从他们上一次生日后的不同时间段进行的,而那些在刚过完生日就接受访问的学生不太有可能在他们当前的年龄尝试第一次吸烟。因此,这里对这些数据进行分析,实际上假设我们访问时是他们最后一次生日,如果他们第一次尝试吸烟时的年龄就是当前的年龄,则将第一次尝试的年龄看做是右删失。

对应于可能的开始吸烟的年龄范围,$\tau_{s-1} \leq T < \tau_s$, $s=1,\cdots,6$,存在6种可能的离散时间存续,$y=s$:

- $s=1$:小于11岁,$\tau_0=0$,$\tau_1=11$。
- $s=2$:11岁,$\tau_1=11$,$\tau_2=12$。
- $s=3$:12岁,$\tau_2=12$,$\tau_3=13$。
- $s=4$:13岁,$\tau_3=13$,$\tau_4=14$。
- $s=5$:14岁,$\tau_4=14$,$\tau_5=15$。
- $s=6$:15岁或以上,$\tau_5=15$,$\tau_6=\infty$(删失)。

注意$s=6$表示总是删失,因为15岁是最大的年纪,并且在当前年龄开始吸烟被视作是右删失的。其建立过程见表12.3。

我们认为的可能影响第一次吸烟年龄的解释变量如下:

- [Girl]虚拟变量,女孩。
- [Boy]虚拟变量,男孩。
- [Cohort2] 虚拟变量,第二次调查的同期群。
- [Parsmoke] 虚拟变量,家中是否有吸烟的父母。

表 12.3 年龄,可能开始吸烟的时间及其关联概率

(a)比例比率

| 年龄 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|
| 可能开始吸烟的时间 | $T<11$ | $T<11$ | $T<11$ | $T<11$ | $T<11$ |
| | $(P_1)$ | $(P_1)$ | $(P_1)$ | $(P_1)$ | $(P_1)$ |
| | $11\leqslant T$ | $11\leqslant T<12$ | $11\leqslant T<12$ | $11\leqslant T<12$ | $11\leqslant T<12$ |
| | $(1-P_1)$ | $(P_2-P_1)$ | $(P_2-P_1)$ | $(P_2-P_1)$ | $(P_2-P_1)$ |
| | — | $12\leqslant T$ | $12\leqslant T<13$ | $12\leqslant T<13$ | $12\leqslant T<13$ |
| | — | $(1-P_2)$ | $(P_3-P_2)$ | $(P_3-P_2)$ | $(P_3-P_2)$ |
| | — | — | $13\leqslant T$ | $13\leqslant T<14$ | $13\leqslant T<14$ |
| | — | — | $(1-P_3)$ | $(P_4-P_3)$ | $(P_4-P_3)$ |
| | — | — | — | $14\leqslant T$ | $14\leqslant T<15$ |
| | — | — | — | $(1-P_4)$ | $(P_5-P_4)$ |
| | — | — | — | — | $15\leqslant T$ |
| | — | — | — | — | $(1-P_5)$ |

(b)当前状态

| 年龄 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|
| | $T<11$ | $T<12$ | $T<13$ | $T<14$ | $T<15$ |
| | $(P_1)$ | $(P_2)$ | $(P_3)$ | $(P_4)$ | $(P_5)$ |
| | $11\leqslant T$ | $12\leqslant T$ | $13\leqslant T$ | $14\leqslant T$ | $15\leqslant T$ |
| | $(1-P_1)$ | $(1-P_2)$ | $(1-P_3)$ | $(1-P_4)$ | $(1-P_5)$ |

资料来源:拉贝-赫斯基思等(Rabe-Hesketh et al., 2001d)。

## 12.3.2 随机截距模型

这里考察两种离散生存次数模型(discrete survival times)、离散时间风险模型(discrete time hazard models)的建模方法,以及(包括删失数据的)定序数据的比例比率模型(proportional odds model)。

可以像2.5.2节的展示2.2那样,对数据进行扩展,从而来估计离散时间风险模型。在扩展后的数据中,处在第一次尝试吸烟的年龄的孩子,在第 $s$ 个间隔的响应为 $y_{ir}$, $r=1,\cdots,s$;在第 $s$ 个间隔期间受访,但第一次吸烟的年龄并不是在他上一次生日之前的孩子,响应值为 $y_{ir}$, $r=1,\cdots,s-1$。如果尝试吸烟发生在相应的间隔(interval),则响应值为1,否则为0。

扩展数据的对数概率回归相当于一个连续比值模型,而余重对数回归则相当于一个比例风险模型(见2.5.2节)。除了常规的协变量外,对每一个可能的第一次尝试间距,模型还包括常量 $\kappa_s$, $s=1,\cdots,5$。

为了对不同学校之间未观测到的异质性进行建模,上述每一个模型的线性估计量都可包括一个随机截距。从而,对于第 $j$ 个学校的第 $i$ 个学生,其第 $r$ 个响应的线性估计量则为

$$\nu_{ijr} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \kappa_r + \zeta_j \qquad (12.1)$$

其中,$\zeta_j \sim \mathrm{N}(0,\psi)$。

同时,还可考虑删失的包含随机截距的比例比率模型。响应概率显示在表12.3中,

其累积概率为

$$P_s \equiv \Pr(T_{ij} < \tau_s) = \frac{\exp(\mathbf{x}'_{ij}\boldsymbol{\beta} + \zeta_j + \kappa_s)}{1 + \exp(\mathbf{x}'_{ij}\boldsymbol{\beta} + \zeta_j + \kappa_s)} = g^{-1}(\mathbf{x}'_{ij}\boldsymbol{\beta} + \zeta_j + \kappa_s)$$

其中，$g$ 是对数概率链接函数。注意，与通常的比例比率模型相反，回归系数 $\beta$ 的符号是逆的，这样一来，正的系数意味着低响应相对于高响应的比率的增加。

对于非删失的存续数据，响应概率具有如下形式：

$$\Pr(\tau_{s-1} \leqslant T < \tau_s) = g^{-1}(\mathbf{x}'_{ij}\boldsymbol{\beta} + \zeta_j + \kappa_s) - g^{-1}(\mathbf{x}'_{ij}\boldsymbol{\beta} + \zeta_j + \kappa_{s-1})$$

因此，可以使用在 2.3.5 节中讨论过的复合链接函数来对模型进行估计，也可以把当前不同年龄孩子的响应，视为含不同分类编号（见表 12.3(a)）差异较大的定序响应，这样来对模型进行估计。例如，调查时年龄为 11 岁的孩子，就有一个包含两个可能分类的一个定序响应，而 12 岁的孩子则有一个包含 3 个分类的响应，以此类推。然后，将阀值 $\kappa_s$ 约束为在不同响应间相等。

表 12.4 的前 3 列中给出了参数估计。所有 3 个模型实际上推导出同样的结论。如果家里没有父母吸烟，女性学生比男性学生更不易处于吸烟的风险中。对于男孩来讲，父母有一个在家中吸烟，就增加了其吸烟的风险，这个效应要大于对女孩的效应。对女孩来讲，存在一个同期群效应，即随着时间的增加，女孩吸第一支烟的风险也在增加。但对男孩来说，似乎不存在这样的同期群效应。尽管所估计的 $\psi$ 非常小，但不同学校间学生开始吸烟的年龄的差异性比较显著（对于连续比例模型和成比例风险模型来说，根据 $\chi^2(1)$ 分布，使用对半 $p$ 值所得到的似然比值检验为 $p<0.001$，见 8.3.4 节）。

**表 12.4　随机截距离散存续模型的估计值和标准误**

| | 连续比值 | | 比例风险 | | 比例比率 | | 当前状态 | | 伸缩效应 | |
|---|---|---|---|---|---|---|---|---|---|---|
| 参　数 | Est | (SE) | Est | (SE) | Est | (SE) | Est | (SE) | Est | (SE) |
| 固定效应： | | | | | | | | | | |
| [Girl] | -0.34 | (0.12) | -0.30 | (0.11) | -0.45 | (0.14) | -0.32 | (0.08) | -0.31 | (0.15) |
| [Parsmoke] | 0.31 | (0.07) | 0.28 | (0.06) | 0.37 | (0.08) | 0.35 | (0.08) | 0.38 | (0.08) |
| [Girl] × [Parsmoke] | 0.29 | (0.09) | 0.26 | (0.09) | 0.33 | (0.11) | 0.38 | (0.12) | 0.36 | (0.11) |
| [Boy] × [Cohort2] | 0.02 | (0.02) | 0.02 | (0.02) | 0.02 | (0.02) | 0.04 | (0.03) | 0.03 | (0.03) |
| [Girl] × [Cohort2] | 0.08 | (0.02) | 0.07 | (0.02) | 0.08 | (0.02) | 0.09 | (0.03) | 0.08 | (0.03) |
| 伸缩效应： | | | | | | | | | | |
| [Boy] | | | | | | | | | -0.10 | (0.04) |
| [Girl] | | | | | | | | | 0.04 | (0.04) |
| 阀值或常数： | | | | | | | | | | |
| $\kappa_1$ | -2.17 | (0.10) | -2.22 | (0.09) | -2.18 | (0.11) | -2.43 | (0.20) | -2.37 | (0.18) |
| $\kappa_2$ | -2.41 | (0.10) | -2.45 | (0.09) | -1.52 | (0.11) | -1.72 | (0.17) | -1.68 | (0.16) |
| $\kappa_3$ | -1.91 | (0.10) | -1.98 | (0.09) | -0.89 | (0.11) | -0.91 | (0.15) | -1.01 | (0.14) |
| $\kappa_4$ | -1.38 | (0.10) | -1.51 | (0.09) | -0.22 | (0.10) | -0.30 | (0.13) | -0.31 | (0.12) |
| $\kappa_5$ | -1.23 | (0.12) | -1.38 | (0.10) | 0.32 | (0.11) | 0.19 | (0.12) | 0.28 | (0.12) |
| 随机效应： | | | | | | | | | | |
| $\psi$ | 0.07 | (0.02) | 0.06 | (0.02) | 0.09 | (0.03) | 0.10 | (0.03) | 0.09 | (0.03) |
| 对数似然值 | -6 225.5 | | -6 225.8 | | -6 223.4 | | -3 487.7 | | -5 948.8 | |

资料来源：拉贝-赫斯基思等（Rabe-Hesketh et al., 2001d）。

### 12.3.3 对"伸缩(telescoping)"效应建模

上面考察的3个模型的一个问题是,它们假设回忆出的第一次吸烟的年龄是可靠的。对开始年龄的测量误差的说明在文献中并不引人关注,考察了下面两种方法:

第一种方法是不考虑孩子响应的时间因素,采用简单对数概率回归模型,把当前吸烟状况指标作为响应变量,仅仅对他们当前的状况(试验过的)建模,作为它们当前年龄的函数,如表12.3(b)所示,表12.4的第4列给出了结果,与成比例比率模型的结果相差无几。

另一种方法是对回忆误差进行建模。有人建议,回忆误差有这样一个显著的特征,它具有从一个较远的过去,到数据收集完成的时间的显著转换(Sudman and Bradburn, 1973;Hobcraft and Murphy, 1987)。这个"伸缩"来源于时间尺度的内在压缩,这样,发生在一个时间 $t$ 之前的事件被认定为发生在时间 $\gamma t$ 之前,$0<\gamma<1$。

伸缩效应也可能来自异方差测量误差,即使误差是对称性分布时,误差方差都是随着事件和回忆的时间之间的落差增加而增加。这是因为,与发生在较近的过去的事件相比,更多的来自于较远过去的、回忆误差较大的事件,被转换到较近的过去;而发生在较近的过去,回忆误差较小的事件,被转换到较远的过去(Pickles et al., 1996)。皮科尔斯等发展了用以区分这些进程的模型,这里只考察一下系统伸缩。

在成比例比率模型中,假设回忆起来的事件开始年龄在给定年龄 $\tau_s$ 之前,此对数比率,对从事件开始年龄以来的时间 $a_{ij}-\tau_s$,是线性减少的,其中 $a_{ij}$是孩子的当前年龄,则

$$\ln\frac{\Pr(T_{ij}<\tau_s)}{1-\Pr(T_{ij}<\tau_s)}=\mathbf{x}'_{ij}\boldsymbol{\beta}+\zeta_j-(a_{ij}-\tau_s)\mathbf{w}'_{ij}\boldsymbol{\alpha}+\kappa_s$$

在这里,$\boldsymbol{\alpha}$ 是系数向量,$\mathbf{w}_{ij}$是能估计伸缩效应程度的解释性变量的向量(系数为正,则说明时间尺度进行了压缩)。如果成比例比率模型被解释为潜响应模型,伸缩效应相当于使阀值取决于事件实际发生时间和回忆起的时间之间的时滞,比如,阀值为

$$-(a_{ij}-\tau_s)\mathbf{w}'_{ij}\boldsymbol{\alpha}+\kappa_s$$

假设伸缩效应的程度只取决于性别,且参数估计值在表12.4中的最后一列显示。这样,对女孩来讲,没有证据表明存在伸缩效应,而对男孩来说则倾向于延伸时间尺度(而非缩短),可能是为了"炫耀"而提前报告首次尝试时间。

值得注意的是,单独的伸缩效应和同期群效应的模型辨识可能就是这样,因为孩子们的部分群体在不同"当前"年龄的两次调查中可能同时存在,因而有不同的时滞。

## 12.4 锻炼和心绞痛:成比例风险随机效应和因子模型

### 12.4.1 介 绍

这里分析了德纳等(Danahy et al., 1976)公布的数据集[1],以及之前皮科尔斯和克劳奇莱(Pickles and Crouchley, 1994, 1995)的考察。冠状心脏疾病受试者参加了随机交叉试验,试验旨在将药物硝异梨醇(ISDN)与安慰剂(placebo)的效果进行比较。

在接到药物(或安慰剂)之前,受试者被要求在自行车上进行锻炼,如果心绞痛

1 数据可以从 gllamm.org/books 上下载。

(angina pectoris)没有发作,就练到力竭。锻炼的时间和结果(发病或者精疲力竭)将被记录下来。然后口头服用药物(或者安慰剂),并且锻炼要在服药后 1,3,5 h 后重复进行。这样,在服用两种药物之前和之后,实际上是对每个受试者重复了"生存"次数。

每名受试者在服用安慰剂条件下重复锻炼测试 4 次,在服用药物条件下重复锻炼测试 4 次。较为关注的响应是在服药状况 $y_{ij}$条件下心绞痛或锻炼到力竭的存续,锻炼测试为 $i$,受试者为 $j$。$d_{ij}$是一个删失指标,心绞痛发生,取值为 1;其他时候,取值为 0。在相应的服用安慰剂条件下,心绞痛 $t_{ij}$存续被视为时随时点变化的变量(在服用安慰剂条件下没有删失)。可惜的是,德纳等(Danahy et al., 1976)中并没有关于提供安慰剂或者药物用药顺序的说明。

因为这 8 次锻炼都是在受试者彻底休息后才开始测试的,所以可以假设导致心绞痛或者锻炼力竭的那个过程,是从每个锻炼测试开始后才启动的,进而,我们假设测试的风险函数对测试是成比例的。根据 12.2 节建构风险组时的几个方面的分类,进行如下设定:

**风险间隔公式**:有时间刻度的间隙时间,在每个锻炼测试开始时从 0 开始。

**基准风险**:每个锻炼测试的共同基准风险。

**第 $k$ 个事件的风险组**:不受约束,因为每个事件发生在单独的锻炼测试中。

这些设定相当于假设了测试之间的成比例风险,以及服用药物之前和之后的成比例的风险。这在设定试验效应为**风险比值**(hazard ratio)时也是必需的。

考察一下下列协变量:

- [Bypass]虚拟变量,接受过冠状动脉支路架桥手术

$$x_{Bj} = \begin{cases} 1 & \text{当进行冠状动脉支路架桥手术时} \\ 0 & \text{其他} \end{cases}$$

- [TimeP]在服用安慰剂条件下发生心绞痛的标准化时间;$t_{ij}$。
- [After]虚拟变量,是否服药,即

$$x_{Tij} = \begin{cases} 0 & \text{当} \quad i = 1 \\ 1 & \text{当} \quad i = 2,3,4 \end{cases}$$

- [Lin]药物效应的线性趋势

$$x_{Dij} = \begin{cases} 0 & \text{当} \quad i = 1 \\ i-3 & \text{当} \quad i = 2,3,4 \end{cases}$$

在每次锻炼测试中发生变化的协变量[TimeP]、[After]和[Lin]的定义汇总在表 12.5中。

**表 12.5 在锻炼测试 $i$ 变动的协变量的定义**

| 测试 $i$ | [TimeP] 安慰剂 $t_{ij}$ | [After] $x_{Tij}$ | [Lin] $x_{Dij}$ |
|---|---|---|---|
| 1 | $t_{1j}$ | 0 | 0 |
| 2 | $t_{2j}$ | 1 | -1 |
| 3 | $t_{3j}$ | 1 | 0 |
| 4 | $t_{4j}$ | 1 | 1 |

## 12.4.2 考克斯回归

考克斯比例风险模型可以表达为(另见式(2.22))

$$h_{ij}(t) = h^0(t)\exp(\nu_{ij})$$

其中,$h^0(t)$是基准风险,即当 $\nu_{ij}=0$ 时的风险。线性估计量为

$$\nu_{ij} = \beta_1 x_{Bj} + \beta_2 t_{ij} + \beta_3 x_{Tij} + \beta_4 x_{Dij} \quad (12.2)$$

其中,$\beta_1$ 表示[Bypass]的效应;$\beta_2$ 表示在安慰剂条件下,发生心绞痛的时间[TimeP]的效应;$\beta_3$ 表示中间的试验后效应,$i=3$,与基准试验效应,$i=1$[After],之间的差异;$\beta_4$ 表示 3 个试验测试之后的效应的线性改变,即[Lin]。

在 2.5.1 节中,在泊松回归中每个风险组包含一个虚拟变量,并在失败次数之间使用对数间隔长度作为偏置,这会导致似然值与偏似然成正比。这种方法的一个问题是,经常有大量的风险组,因而,大量的参数要估计。比如,对于现在的小数据组来说,需要多达 64 个虚拟变量。这意味着以时间的精简平滑函数(parsimonious smooth function)来近似于基准风险。对此,我们考察一个 6 阶的正交多项式(orthogonal polynomial):

$$\ln h^0(t) = \alpha_0 + \alpha_1 f_1(t) + \alpha_2 f_2(t) + \alpha_3 f_3(t) + \alpha_4 f_4(t) + \alpha_5 f_5(t) + \alpha_6 f_6(t) \quad (12.3)$$

其中,$f_k(t)$是第 $k$ 个项。尽管基准风险建模为时间的平滑函数,分段指数方程(piecewise exponential formulation)表明风险是连续事件之间的一个常数,其中各阶段的高度由平滑函数决定。

[Bypass]、[TimeP]、[After]、[Lin]的效应的估计值和标准误差如表 12.6 所示,它是偏似然(考克斯回归)、含风险组虚拟变量的最大似然泊松分布及包含时间的 6 阶正交多项式(orthpol-6)的最大似然泊松回归的列联表。对每个不同的回归过程,我们报告所估计的$\beta'$值,标准误差(SE),以及风险比值 $\exp(\beta)$(标记为 HR)。理论上说,偏似然回归模型和含虚拟变量的最大似然泊松回归模型的估计值和标准误差完全相同。另外请注意,通过将对数基准风险作为时间的 6 阶正交多项式来建模,也可得到与考克斯回归几乎相同的估计值。这个方法比在每个风险组中都包含一个虚拟变量的泊松回归更为简约,毕竟在当前的例子中,那样将需要 57 个额外的参数。对于更大的数据集来说,这么大量的额外参数很难处理,因而,在后续的潜变量存续模型中,我们将主要使用这个对数基准风险 $\ln[h^0(t)]$的正交多项式近似值。

**表 12.6　几种考克斯回归模型的估计值**

| | 偏似然 | | | 最大似然泊松虚拟变量 | | | 最大似然泊松 orthpol-6 | | |
|---|---|---|---|---|---|---|---|---|---|
| 参　数 | Est | (SE) | HR | Est | (SE) | HR | Est | (SE) | HR |
| $\beta_1$[Bypass] | 0.88 | (0.29) | 2.42 | 0.88 | (0.29) | 2.42 | 0.90 | (0.30) | 2.45 |
| $\beta_2$[TimeP] | −1.17 | (0.20) | 0.31 | −1.17 | (0.20) | 0.31 | −1.12 | (0.20) | 0.33 |
| $\beta_3$[After] | −1.02 | (0.29) | 0.36 | −1.02 | (0.29) | 0.36 | −1.00 | (0.29) | 0.37 |
| $\beta_4$[Lin] | 0.62 | (0.19) | 1.85 | 0.62 | (0.19) | 1.85 | 0.59 | (0.19) | 1.81 |
| 对数似然值 | | −224.76 | | | −284.14 | | | −320.22 | |

图 12.2 显示了含风险数据组虚拟变量的泊松模型中,估计常数及其 95% 置信区间。平滑曲线是泊松模型的估计对数基准风险,泊松模型含一个时间的 6 阶多项式。多项式看起来并没有使对数基准风险过于平滑,毕竟曲线落在了常数的置信区间内。

### 12.4.3　多元正态潜变量的比例风险回归

首先考察传统的**脆弱模型**,即含线性估计量的简单随机截距模型。

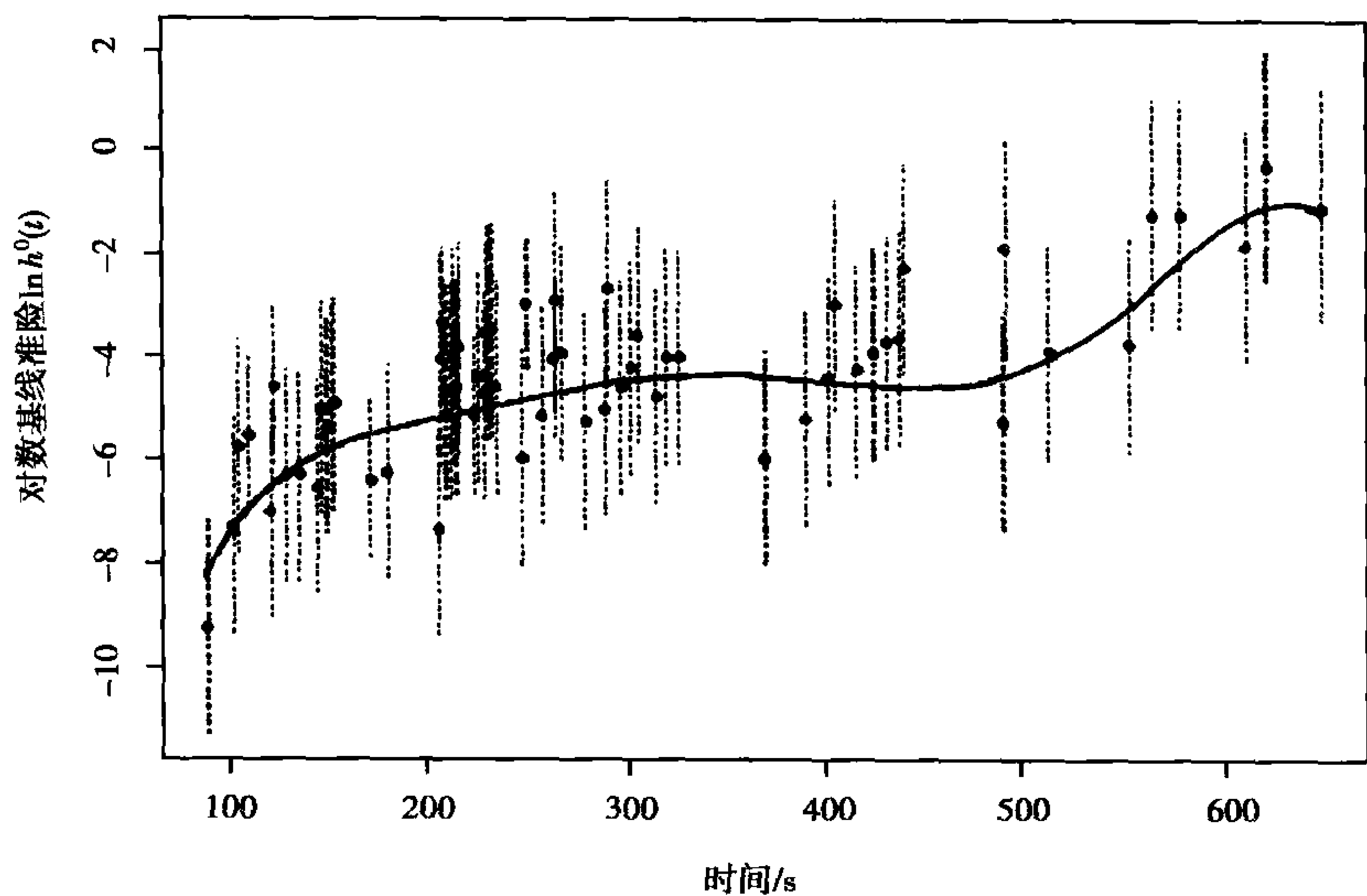



图 12.2 心绞痛数据的对数基准风险。其中,图中的点表示置信区间为 95% 的每个风险组的估计常数,曲线表示相应的 6 阶多项式

$$\nu_{ij} = \beta_1 x_{Bj} + \beta_2 t_{ij} + \beta_3 x_{Tij} + \beta_4 x_{Dij} + \zeta_{1j}$$

其中,$\zeta_{1j} \sim N(0,\psi_{11})$。随机截距的指数 $\exp(\zeta_{1j})$,经常被称为**脆弱性**,因此设定了一个对数正态脆弱(请参见 McGilchrist and Aisbett, 1991),其选项分布包括单参数(Gamma)(请参见 Clayton,1978)、正稳定(请参见 Hougaard, 1986b)和逆高斯(请参见 Hougaard, 1986a)。

随机截距模型的估计值和标准误差在表 12.7 的第 2 列中。确定部分的估计值也在这个指数表中,如风险比值 HR。

**表 12.7 随机截距(脆弱)估计值和因子结构的比例风险模型**

| | 脆弱模型 | | | 因子模型 | | |
|---|---|---|---|---|---|---|
| 参　数 | Est | (SE) | HR | Est | (SE) | HR |
| 确定部分: | | | | | | |
| $\beta_1$[Bypass] | 1.09 | (0.66) | 2.96 | -0.03 | (0.60) | 0.97 |
| $\beta_2$[TimeP] | -1.44 | (0.30) | 0.24 | -1.86 | (0.34) | 0.16 |
| $\beta_3$[After] | -0.92 | (0.31) | 0.40 | -1.41 | (0.51) | 0.24 |
| $\beta_4$[Lin] | 0.81 | (0.21) | 2.24 | 1.14 | (0.33) | 3.11 |
| 随机部分: | | | | | | |
| $\psi_{11}$ | 1.32 | (0.69) | | 8.86 | (4.97) | |
| $\lambda_1$ | | | | 0.19 | (0.10) | |
| $\lambda_2$ | | | | 1 | — | |
| $\lambda_3$ | | | | 0.71 | (0.14) | |
| $\lambda_4$ | | | | 0.50 | (0.14) | |
| 对数似然值 | | -311.19 | | | -301.41 | |

为了允许约束更少的依赖结构,我们放弃传统的脆弱模型,而考察以下**因子模型**:

$$\nu_{ij} = \beta_1 x_{Bj} + \beta_2 t_{ij} + \beta_3 x_{Tij} + \beta_4 x_{Dij} + \lambda_i \zeta_{1j} \qquad (12.4)$$

这里,$\zeta_{1j} \sim N(0,\psi_{11})$,且已经通过约束一个因子负荷对模型进行了辨识,$\lambda_2 = 1$。如果因子模型中,$\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = 1$,就可得到脆弱模型。因子模型的估计值和标准误差

在表 12.7 中的第 3 列。采用这个模型比随机截距模型更好些，因为对数似然值增加很多。

另一个可选择的模型是**随机试验模型**(random treatment model)，其中[After]有一个关联的随机系数 $\zeta_{2j}$，但没有随机截距，即

$$\nu_{ij} = \beta_1 x_{Bj} + \beta_2 t_{ij} + (\beta_3 + \zeta_{2j}) x_{Tij} + \beta_3 x_{Dij} \tag{12.5}$$

其中，$\zeta_{2j} \sim N(0, \psi_{22})$。随机试验模型的估计值和置信区间在表 12.8 的第 2 列中。这个模型是因子模型的一个特例，即 $\lambda_1 = 0$ 且 $\lambda_2 = \lambda_3 = \lambda_4 = 1$。加入几个约束值后，对数似然的减少有限，这也可以通过检查因子负荷推测出来。图 12.3 显示了第 3 个锻炼测试的对数风险，[TimeP]等于均值(240.5)，对数脆弱为 -2，-1，0，1，2 标准差。

**表 12.8 随机系数比例风险模型的估计值**

| | 随机试验 | | | 随机截距和试验 | | |
|---|---|---|---|---|---|---|
| 参　数 | Est | (SE) | HR | Est | (SE) | HR |
| 固定部分: | | | | | | |
| $\beta_1$[Bypass] | 0.58 | (0.44) | 1.78 | 0.47 | (0.54) | 1.61 |
| $\beta_2$[TimeP] | -1.85 | (0.32) | 0.15 | -1.75 | (0.33) | 0.17 |
| $\beta_3$[After] | -1.09 | (0.46) | 0.34 | -1.16 | (0.42) | 0.32 |
| $\beta_4$[Lin] | 0.78 | (0.21) | 2.18 | 0.83 | (0.22) | 2.30 |
| 随机部分: | | | | | | |
| $\psi_{11}$ | | | | 0.20 | (0.30) | |
| $\psi_{22}$ | 2.62 | (1.24) | | 1.74 | (1.09) | |
| $\psi_{21}$ | | | | †0.59 | (0.42) | |
| 对数似然值 | | -305.48 | | | -304.53 | |

注：†边界解。

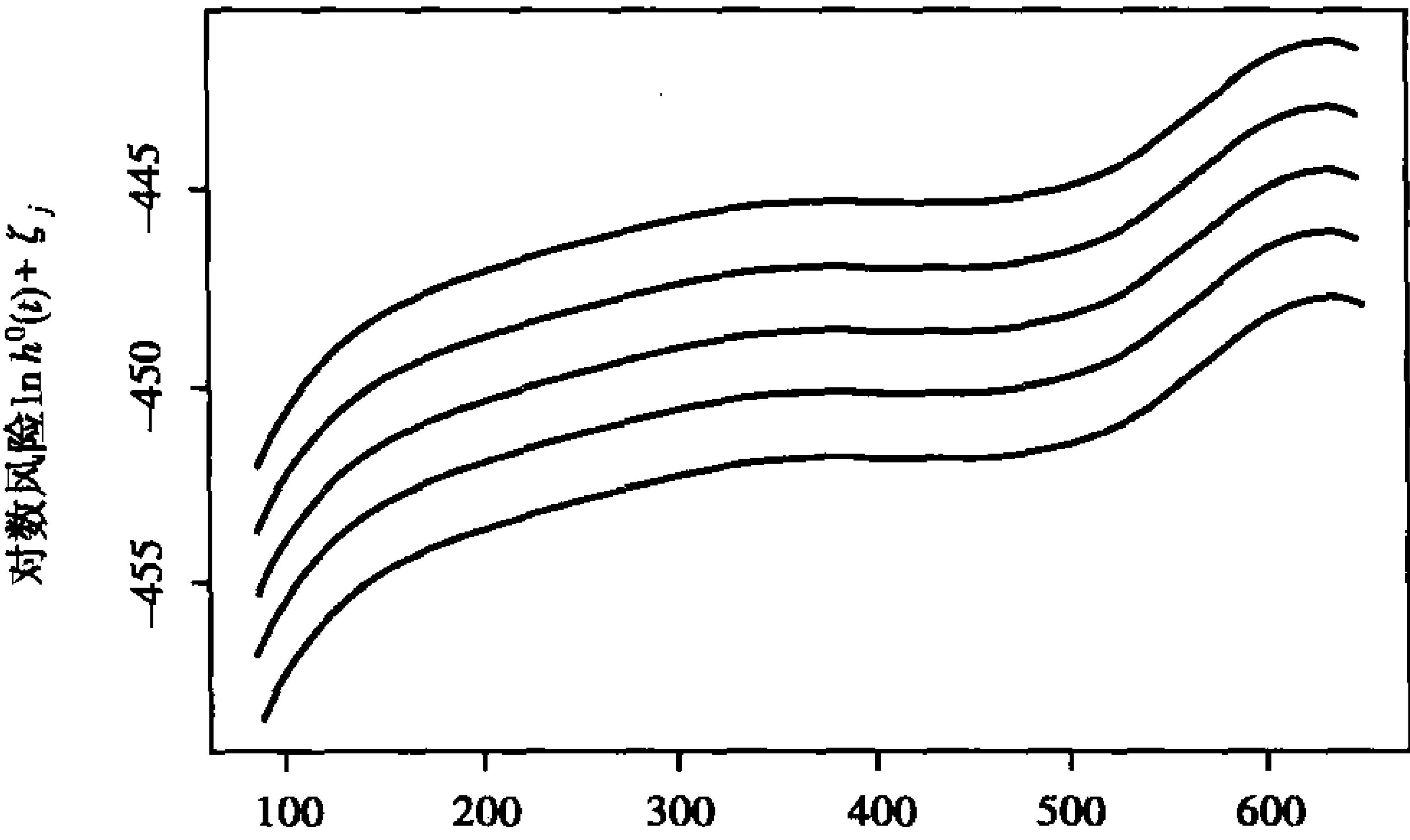



**图 12.3 假象对象在训练测试 3 的对数风险。**

**对数脆弱 $\zeta_j$ 等于 -2，-1，0，1 和 2，所估计的标准差 $\hat{\psi}^{1/2}$**

最终，考察一个**随机截距试验模型**。这个模型有两个随机效应，随机截距 $\zeta_{1j}$ 与随机试验效应 $\zeta_{2j}$ 交互相关：

$$\nu_{ij} = \zeta_{1j} + \beta_1 x_{Bj} + \beta_2 t_{ij} + (\beta_3 + \zeta_{2i}) x_{Tij} + \beta_4 x_{Dij}$$

其中,$(\zeta_{1j},\zeta_{2j})$假设为二元正态分布,方差为$\psi_{11}$,$\psi_{22}$,协方差为$\psi_{21}$。

多变量正态分布对含多个潜变量的模型来说是一种方便的选择,因为如何归纳另外的(单变量)常用于脆弱模型的分布并不容易发现。模型的估计值和标准误差在表12.8的第2列中。实际上,随机效应之间的相关值为1,而且边界解也表明一个随机效应能满足应用,这也能通过与随机试验模型比较中,对数似然值有一个小的增长反映出来。因此,就放弃了随机截距和试验模型。

在表12.7和12.8中固定部分的估计风险比值HR对潜变量条件依赖。除了因子模型外,对于所有模型来说,[Bypass]风险比值的估计值要大于1。尽管如此,估计值很不精确,并且在不同模型中变化很大,从2.96到0.96。正如所期望的,在服用安慰剂条件下,较长的存续似乎减少了在药物情况下的风险。[After]风险比值的估计值表明,与基准风险相比,药物减少了试验中期、后期锻炼测试的风险。同样,对数风险的增长超过了试验后的测试,因为$\beta_4$估计为正数。

至于随机部分,在随机截距模型中的随机截距方差似然比检验表明,它仍是必需的,尽管与所对应的瓦尔德检验相矛盾。建议进行对半$p$值似然比测试,以检测潜变量的方差为零假设,见8.3.4节。

### 12.4.4 潜类比例风险模型

在某些情况下需要假设随机效应为其离散,即潜类模型。基于连续随机效应的结果,在本节中,我们主要关注随机试验和因子模型。模型与连续情形的例子有相同的公式,其关键的差别在于随机效应$\zeta_j$为离散的,即

$$\zeta_j = e_c, c = 1,\cdots,C$$

且概率为$\pi_C$,而非正态分布。这里只讨论双分类模型$C=2$。

在式(12.5)中,双分类随机试验模型的估计值在表12.9的第2列。随机试验模型第1个类别的原点-2.59表明,有一类病人对药物回应得极好。这一类包含了全体的21%。

对于固定部分的估计值,两个潜类模型的解与表12.7中的正态随机试验和因子模型的估计值相似。就假设模型为正态这点来讲,因子模型的负荷的模式是相同的。

**表12.9 双分类随机试验模型和因子模型的估计值**

| | 随机试验模型 | | | 因子模型 | | |
|---|---|---|---|---|---|---|
| 参　数 | Est | (SE) | HR | Est | (SE) | HR |
| 固定部分: | | | | | | |
| $\beta_1$[Bypass] | 0.28 | (0.29) | 1.32 | 0.28 | (0.30) | 1.33 |
| $\beta_2$[TimeP] | -1.95 | (0.26) | 0.14 | -1.97 | (0.26) | 0.14 |
| $\beta_3$[After] | -1.22 | (0.41) | 0.30 | -1.39 | (0.47) | 0.25 |
| $\beta_4$[Lin] | 0.85 | (0.20) | 2.33 | 1.13 | (0.34) | 3.08 |
| 随机部分: | | | | | | |
| Var($\eta_j$) | 1.75 | | | 5.52 | | |
| $\lambda_1$ | | | | 0.07 | (0.10) | |

续表

| 参 数 | 随机试验模型 | | | 因子模型 | | |
|---|---|---|---|---|---|---|
| | Est | (SE) | HR | Est | (SE) | HR |
| 随机部分: | | | | | | |
| $\lambda_2$ | | | | 1 | — | |
| $\lambda_3$ | | | | 0.68 | (0.29) | |
| $\lambda_4$ | | | | 0.40 | (0.20) | |
| 原点: | | | | | | |
| $e_1$ | -2.59 | | | -4.63 | | |
| $e_2$ | 0.68 | | | 1.19 | | |
| 类别概率: | | | | | | |
| $\Pr(e_1)$ | 0.21 | | | 0.21 | | |
| $\Pr(e_2)$ | 0.79 | | | 0.79 | | |
| 对数似然值 | | -303.82 | | | -301.42 | |

### 12.4.5 非参数最大似然估计(Nonparametric maximum likelihood estimates)

随机效应的参数连续性分布,比如前面出现的多变量正态分布,已经在一些研究者中间引发了恐慌。大部分人承认如果模型结果对参数分布的选择高度敏感,则是有问题的,因为参数分布的选择往往出于方便的动机。另外一种选择是采用上面提到的包含潜类别的模型,但是如果假设总体是由真实离散类别组成的,这又看似不合理。实际上,这已经引发了存续模型的发展,此模型中,潜变量分布可能是连续的和/或者离散的,不需要任何先验设定,并且潜变量的原点和质量点能够得以估计。甚至,当通过引入额外的质量点(请参见 Heckman and Singer, 1982, 1984)并不能改善似然值的话,那么就可以得到非参数最大似然估计值(NPMLE),见 6.5 节。

**表 12.10 随机试验模型和因子模型的 NPMLE**

| 参 数 | 随机试验模型 3 个质量点 | | | 因子模型 5 个质量点 | | |
|---|---|---|---|---|---|---|
| | Est | (SE) | HR | Est | (SE) | HR |
| 固定部分: | | | | | | |
| $\beta_1$[Bypass] | 0.33 | (0.30) | 1.39 | -0.22 | (0.36) | 0.81 |
| $\beta_2$[TimeP] | -2.18 | (0.29) | 0.11 | -2.33 | (0.34) | 0.10 |
| $\beta_3$[After] | -1.07 | (0.48) | 0.34 | -2.18 | (1.12) | 0.11 |
| $\beta_4$[Lin] | 0.84 | (0.21) | 2.34 | 1.94 | (0.94) | 6.93 |
| 随机部分: | | | | | | |
| $\mathrm{Var}(\zeta_j)$ | 2.92 | | | 47.24 | | |
| $\lambda_1$ | | | | 0.08 | (0.05) | |
| $\lambda_2$ | | | | 1 | | |
| $\lambda_3$ | | | | 0.54 | (0.11) | |
| $\lambda_4$ | | | | 0.23 | (0.13) | |
| 对数似然值 | | -301.34 | | | -295.59 | |

随机试验模型的 NPML 估计值在表 12. 10 的第 2 列中。所得到的 3 个质量点的解的原点和质量估计值分别是( -2. 91,0. 32,2. 84)和(0. 21,0. 66,0. 14),因子模型(12. 4)中的 NPML 估计值在表 12. 10 的第 3 列中分别给出。所得到的 5 个质量点的解包括原点估计( -18. 10, -5. 51,0. 74,3. 88,6. 91)以及相应的质量(0. 10,0. 10,0. 28,0. 44,0. 08)。要注意,第一类别含一个趋 0 的脆弱,表示分界解。所估计的 Var($\zeta_j$)值较大也反映了这一点。估计值与那些假定是正态分布的因子接近,尽管因子的非参数分布似乎是偏态的。因此,在此例子中,因子模型能够很好地应付因子分布的设定错误。

## 12. 5 总结与进一步阅读

起初,我们讨论了缘于建构多重事件群集存续数据的风险组过程中的特殊问题,指出此类例子中恰当的建模要比单一事件群集存续数据的建模更为复杂。

第一个针对青少年吸烟的起始问题的应用,对青少年开始抽烟进行离散时间脆弱模型分析。探讨了处理回忆初次吸烟年龄的记忆偏差的方法,并且对"系统的伸缩"模型进行了估计。

然后,在一个随机临床试验中,考察了心绞痛发作的比例风险模型。引入不同种类的潜变量,来对每个受试者和重复存续之间进行相关性进行建模。有连续性潜变量的模型,如因子模型,各类的随机效应模型,以及有离散潜变量的模型,都对其进行了估计。同时,采用非参数最大似然的方法,未对潜变量分布进行设定。

休加(Hougaard, 2000)提供了应付多变量脆弱模型的一个扩展的试验,即关注正稳定脆弱分布的模型。沃芒(Vermunt, 1997)也提出了一种容易理解的离散潜变量的生存模型。关于含脆弱和潜变量的生存和存续建模,有诸多有益的论文,包括 Clayton(1978)、Vaupel et al. (1979)、Clayton and Cuzick (1985)、Aalen (1988)、Pickles and Crouchley (1994,1995)、Vaida and Xu(2000)。

近年来,存续和其他响应函数的联合建模引起了此研究的广泛兴趣,可以阅读霍根和莱尔德(Hogan and Laird, 1997ab)作为一个概括了解。我们考察了对肝硬化的生存联合模型,在 14.6 节中有些里程碑式的研究进展,估计了试验的直接和间接效应。

# 13 比较响应：多分类响应及排序数据

## 13.1 介　绍

在2.4.3节中，我们介绍了多分类响应（如“首选项”“离散选项”）、成对比较及排序。对这些过程的潜响应的形式化表达，其显著的特征即为响应是一系列效用的向量。

假设不同选项的效用之间互不相关是不现实的。否则，对于多项式对数概率模型来说，这将潜含一个不现实的假设，即“无关选项独立性”，我们在13.2节再来讨论这个问题。对这个依赖性建模，通常会像多元回归模型那样简单地在模型中包含变量以及随机参数。然而，与连续响应变量的情形相反，我们并未直接观测到效用的多元向量，因为响应过程是通过相互比较进行的。正如邦奇（Bunch, 1991）、齐内（Keane, 1992）以及斯科隆多和拉贝-赫斯基思（Skrondal and Rabe-Hesketh, 2003b）所讨论的，在这种情形下，模型辨识将很困难。在13.3节中，我们将介绍在一般情况下模型的固定部分及随机部分是如何构成的。

在本章，我们分析潜变量的多项式对数概率模型，而非正态概率模型。其理由出于实用性考虑，即这两个模型非常类似，但是对数概率模型在计算上更加方便。具体来说，就是给定潜变量，响应的条件概率可用闭式多项式对数概率模型来表示，而用正态概率模型则不能。在2.4.3节也已指出，在多项式正态概率模型中要得到响应概率需要进行积分计算，即使模型并不包含潜变量。

在本章中所使用的例子来自政治学以及营销学。我们将讨论包含连续及离散潜变量的排序模型以及离散选择模型。

## 13.2 异质性及“无关选项独立性”

根据2.4.3节的对数概率离散选择模型(2.18)，对于第 $i$ 个单元的两个选项 $a$ 和 $b$，其比率为

$$\frac{\Pr(a)}{\Pr(b)} = \exp(\nu_i^a - \nu_i^b) \tag{13.1}$$

该比率依赖于这两个选项的线性估计。因此，该比率并不依赖于在可选集中的其他选项。在排序的情形中，该比率并不依赖于可选集中其他已被选中的选项，已被选中的选项个数，以及选项被选中的顺序。鲁斯（Luce, 1959）将此属性称为“无关选项独立性”（IIA）。

这个IIA属性我们可用麦克法登（McFadden,1973）的红-蓝公共汽车例子来演示。假设存在3个政党，劳工党Ⅰ、劳动党Ⅱ和保守党。前两个政党不太容易区分，且具有相同

的线性估计量 $\nu_i^{\text{lab}}$，而保守党的线性估计量为 $\nu_i^{\text{con}}$。那么投票给劳工党Ⅰ或劳工党Ⅱ的概率为

$$\frac{2\exp(\nu_i^{\text{lab}})}{2\exp(\nu_i^{\text{lab}}) + \exp(\nu_i^{\text{con}})}$$

假设将这两个政党合并成为一个单独的劳工党，那么投票给劳工党的概率降为

$$\frac{\exp(\nu_i^{\text{lab}})}{\exp(\nu_i^{\text{lab}}) + \exp(\nu_i^{\text{con}})}$$

很显然，这是不符合直觉的。该模型限制性过强（如 Takane, 1987 及其参考文献）。对于一个无任何偏向性的投票者来说，IIA 的结果是很清楚的，即 $\nu_i^{\text{lab}} - \nu_i^{\text{con}} = 0$。在这个例子中，为了保证等概率选择，两个政党的合并将导致投票给劳工党的概率从 0.67 降到 0.50。然而，在现实中，所有的投票者都存在党派偏好，从而存在异质性。这种异质性，无论是观测到的还是未观测到的，保证了模型不会潜含因合并而导致劳工党投票率降低的意思。

为了说明观测到的异质性，我们分析一下性别的固定效应。假设对男性，$\nu_i^{\text{lab}} - \nu_i^{\text{con}} = -1.2$；对女性，$\nu_i^{\text{lab}} - \nu_i^{\text{con}} = 2.8$。对于一个男性占 50% 的总体而言，合并将导致投给劳工党的边缘概率从 0.67 降到 0.59，从而也就违背了 IIA。在实践中，观测到的协变量无法解释个体党派偏好的全部变异性。剩余的未观测到的异质性可通过在模型的线性估计量中包含这两个劳工党的公共随机效应。例如，如果对男性，$\nu_i^{\text{lab}} - \nu_i^{\text{con}} = -0.8 + \zeta_i$；对女性，$\nu_i^{\text{lab}} - \nu_i^{\text{con}} = 3.2 + \zeta_i$。其中，$\zeta_i$ 为正态分布的随机截距，方差等于 16，那么劳工党的边缘（关于 $\zeta_i$）概率在合并前为 0.67，在合并后为 0.63。增加固定效应的值或随机效应方差将会进一步缩小这个差距。

如果不包含随机效应，那么，所带来的其中一个结果就是不现实的 IIA 属性仍然对观测到的协变量条件成立。例如，当性别为唯一观测到的协变量时，IIA 存在于两个性别组之中。即使考虑到投票者特定的随机效应，对于给定投票者，IIA 仍然成立。对于这种情形，在模型可以辨识的前提下，可以通过包含投票场景特定的随机效应来消除。

## 13.3 模型结构

我们将对斯科隆多和拉贝-赫斯基思（Skrondal and Rabe-Hesketh, 2003b）所提出的模型结构做个简单介绍，更多内容请参考该书。简单来说，我们考虑一个双层模型，单元 $i$（第一层）嵌套在组 $j$（第二层）中。

对于定类响应的潜效用 $u_{ij}^s$ 的多层次模型可用下面公式表示

$$u_{ij}^s = \nu_{ij}^s + \epsilon_{ij}^s$$

根据 2.4.3 节，当 $\epsilon_{ij}^s$ 被设定为独立的冈贝尔分布时，多项式对数概率模型有解。线性估计量 $u_{ij}^s$ 表示类别或选项 $s$ 的效用均值，并在第一层上含固定部分 $f_{ij}^s$，在第二层上含随机部分 $\delta_{ij}^{s(1)}$ 与 $\delta_{ij}^{s(2)}$，即

$$u_{ij}^s = f_{ij}^s + \delta_{ij}^{s(1)} + \delta_{ij}^{s(2)} + \epsilon_{ij}^s$$

对于多分类响应，通常假设被选中的为效用（偏好）最高的选项；对于排序，则假设选项按效用的高低顺序排序。

现在，我们接下来开始介绍模型的固定部分以及随机部分的结构。

### 13.3.1 固定部分

模型的固定部分被定义为

$$f_{ij}^{s} = m^{s} + \mathbf{x}_{ij}^{s\prime}\mathbf{b} + \mathbf{x}_{ij}^{\prime}\mathbf{g}^{s} \quad (13.2)$$

其中,$m^s$ 为常数项,$\mathbf{x}_{ij}^s$为在选项上变动的向量,同时也可能在单元/群组上变动,向量 $\mathbf{x}_{ij}$在单元/群组上变动,但不在选项上变动。所对应的固定系数向量分别为 $\mathbf{b}$ 与 $\mathbf{g}^s$。请注意,效应 $\mathbf{b}$ 被假设为在所有效用上都相同,以使得模型的该部分表示效用与选项(也可能是单元/群组)特定的协变量之间的线性关系。例如,对于 13.4 节中的投票例子,我们假设第 $i$ 个投票者对第 $s$ 个党派在左-右翼政治维度上的距离的量度与投票者对该党派的效用存在线性关系。对于某些选项及单元/群组特定的变量,其效应可在不同选项之间变动。通过在 $\mathbf{x}_{ij}^s$中包含这些变量的交互项以及选项的虚拟变量可以包容这种效应。

特雷恩(Train, 1986)讨论了仅包含固定效应的离散选择模型,查普曼和史德林(Chapman and Staelin, 1982)以及埃里森和克里斯塔基斯(Allison and Christakis, 1994)则讨论了排序模型。当模型(13.2)中的 $\mathbf{x}_{ij}^{\prime}g^2$ 与 $m^s$ 都被忽略时,就称为计量经济学中标准的条件对数概率模型。条件对数概率模型用于离散选择(如 McFadden, 1973)、排序(如 Hausman and Ruud, 1987)以及成对比较(如 Bradley and Terry, 1952)中。当模型(13.2)中的 $\mathbf{x}_{ij}^{s\prime}\mathbf{b}$ 被忽略时,就是生物统计中用于离散选择的标准模型,也被称为多分类对数概率回归模型(如 Hosmer and Lemeshow, 2000)。

### 13.3.2 第一层次上的随机部分

方程项 $\delta_{ij}^{s(1)}$ 在单元内带来选项间的依赖,其结构为

$$\delta_{ij}^{s(1)} = \mathbf{z}_{ij}^{s(1)\prime}\boldsymbol{\xi}_{ij}^{(1)} + \boldsymbol{\lambda}^{s(1)\prime}\boldsymbol{\eta}_{ij}^{(1)} \quad (13.3)$$

其中,$\boldsymbol{\xi}_{ij}^{(1)}$ 为随机效应,允许选项特定的协变量 $\mathbf{z}_{ij}^{s(1)}$ 的效应在单元 $i$ 中变化,$\boldsymbol{\eta}_{ij}^{(1)}$ 为表示在 $u_{ij}^s$上的效应为 $\boldsymbol{\lambda}^{s(1)}$ 的未观测到的变量因子。或者,$\boldsymbol{\lambda}^{s(1)}$ 可以被解释为选项 $s$ 的未被观测到的属性,$\boldsymbol{\eta}_{ij}^{(1)}$ 可被解释为在效用上的随机效应。在投票例子中,设 $z_{ij}^{s(1)} = x_{ij}^{s(1)}$ 表示左-右翼政治维度的距离,随机斜率 $\xi_{ij}^{(1)}$ 允许政治距离的效应在投票者中变动。

麦克法登和特雷恩(McFadden and Train, 2000)在多项式对数概率模型中讨论了一个上述类型的模型。豪斯曼和怀斯 (Hausman and Wise, 1978)分析了一个只包含随机系数的特殊模型。布洛克瑟姆(Bloxom, 1972) 以及阿巴克尔和纽金特(Arbuckle and Nugent, 1973) 针对成对比较,布拉迪(Brady, 1989)、博肯哈特(Böckenholt, 1993) 以及陈卫和本特勒(Chan and Bentler, 1998)针对排序提出了只包含公共因子的模型。必须一提的是,对于离散选择来说,单元层次的因子模型是无法辨识的,除非单元及选项特定的协变量被包含在固定部分中(如 Keane, 1992)。

### 13.3.3 第二层次的随机部分

方程项 $\delta_{ij}^{s(2)}$,在群组中带来效用间的依赖,其结构为

$$\delta_{ij}^{s(2)} = \mathbf{z}_{ij}^{s(2)\prime}\boldsymbol{\xi}_{j}^{(2)} + \mathbf{z}_{ij}^{(2)\prime}\boldsymbol{\gamma}_{j}^{s(2)} + \boldsymbol{\lambda}^{s(2)\prime}\boldsymbol{\eta}_{j}^{(2)} \quad (13.4)$$

其中,$\boldsymbol{\xi}_{j}^{s(2)}$ 与 $\boldsymbol{\gamma}_{j}^{s(2)}$ 为随机系数向量,其对应的变量向量为 $\mathbf{z}_{ij}^{s(2)}$ 与 $\mathbf{z}_{ij}^{(2)}$。$\boldsymbol{\xi}_{j}^{(2)}$ 允许选项特定的协变量的效应在群组 $j$ 间变动,其中,$\boldsymbol{\gamma}_{j}^{s(2)}$ 表示单元特定的协变量 $\mathbf{z}_{ij}^{(2)}$ 的效应在群组间的变差。请注意,在这个例子中,对于每个选项存在一个随机系数。在投票的这个例子中,$\boldsymbol{\xi}_{j}^{(2)}$ 可表示支持者的政治距离效应的异质性。当 $\mathbf{z}_{ij}^{(2)} = 1$ 时,$\boldsymbol{\gamma}_{j}^{s(2)}$ 为随机截距,表示在

支持者自身内变动的协变量(如年龄)的随机系数(在支持者中间变动)。$\boldsymbol{\eta}_j^{(2)}$ 表示在群组层次上的未观测到的异质性因素,在 $u_{ij}^s$ 上的效应为 $\boldsymbol{\lambda}^{s(2)}$。在第一层次与第二层次上的随机项是类似的,唯一的区别在于第一层次上没有对应于 $\gamma_j^{s(2)}$ 的项,否则将很难辨识模型。

对于单元特定的协变量仅包含随机系数的特殊情形,海德克(Hedeker, 2003)以及丹尼尔和盖特索尼斯(Daniels and Gatsonis, 1997)在离散选择研究中,博肯哈特(Böckenholt, 2001a)在成对比较研究中进行了讨论。利维特和特雷恩(Revelt and Train, 1998)设定了一个特殊模型,其中,选项特定或选项及单元特定的协变量仅包含随机系数。哈埃耶等(Haaijer et al., 1998)在结合选择实验(conjoint choice experiment)中使用了一个在选项特定的协变量上包含随机系数的模型。特殊的例子还包括艾洛德和齐内(Elrod and Keane, 1995)的仅有公共因子的模型。麦克法登和特雷恩(McFadden and Train, 2000)以及其他一些学者提出了既包含随机系数也包含因子结构的"混合对数概率模型(mixed logit model)"。博肯哈特(Böckenholt, 2001b)将该模型应用于排序研究中,另外还包含了离散随机截距 $\gamma_j^{s(2)}=e_{jc}^s, c=1,\cdots,C$。

仅包含离散随机截距的潜类模型由鎌仓等(Kamakura et al., 1996)应用在离散选择数据的分析上,克洛恩(Croon, 1989)将其应用在排序研究上,迪伦等(Dillon et al., 1993)以及其他学者将其应用在成对比较研究上。对于潜类模型在排序及离散选择中的应用,我们将分别在13.5节和13.6节中给出例子。

据我们所知,同时在单元及群组层次上对依赖性建模这个问题作出贡献的仅有阿伦比和伦克(Allenby and Lenk, 1994)以及斯科隆多和拉贝-赫斯基思(Skrondal and Rabe-Hesketh, 2003b),尽管前者并未给出表示单元层次异质性的明显定义。

## 13.4 不列颠普选:离散选择及排序的多层次模型

斯科隆多和拉贝-赫斯基思(Skrondal and Rabe-Hesketh, 2003ab)分析了不列颠选举研究1987—1992年的面板数据(Heath et al., 1991, 1993, 1994)[1]。数据共含1 608个受访者。协变量上缺失的投票以及并未投给主要党派候选人的数据被排除在外。最后形成的数据包括2 458个投票记录,1 344位投票者以及249个选区。

选项有保守党、劳工党和自由党(联盟),分别标记为con,lab,lib,并分别对应于 $s=1,2,3$。投票者并未被直接要求对候选人进行排序,但是很明显,第一个被选择的候选人对应于排序1。另外,投票者也对各党派进行了从"非常反对"到"非常支持"的5级评分。我们将该评分作为排序值赋给其余的候选人,从而可以将候选人按评分排序。

要考虑的协变量如下:

- [LRdist] 投票者的政治态度在左-右翼政治维度上的位置,与其所支持的党派在左-右翼政治维度位置的距离。为了避免"合理性问题(rationalization problems)"(如Brody and Page, 1972),使用根据全体投票者计算的党派均值。距离的量度通过4个问题建构,受访者将他们自己以及每个党派在一个11级评分表上进行评分,而每个问题都是一组对立的句子(当前重要的是处理失业问题还是通过膨胀;增进政府服务还是减税;国有化还是私有化;在重新分配财富上应该投入更多还是应该减少投入)。

---

1 数据由英国数据中心提供,本分析所用的数据子集可从 gllamm.org/books 下载(courtesy of Anthony Heath)。

- [1987] 虚拟变量,1987 年全国竞选。
- [1992] 虚拟变量,1992 年全国竞选。
- [Male] 虚拟变量,投票者性别为男性。
- [Age] 投票者的年龄,以 10 年为单位。
- [Manual] 虚拟变量,投票者的父亲是否从事手工业。
- [Inflation] 自从上次选举以来对通货膨胀程度的主观 5 级评分。

该数据具有层次结构,投票 $i$(第一层次)嵌套在投票者 $j$(第二层次),投票者嵌套在选区 $k$(第三层次)之中。变量[LRdist]是可选的,由于其对于每次选举都是不同的,因而用 $x^s_{ijk}$ 表示。变量[1987]、[1992] 及[Inflation]是每次选举特定的,标记为 $\mathbf{x}_{ijk}$,变量[Male]、[Age]及[Manual]是投票者特定的,标记为 $\mathbf{x}_{jk}$。

模型的固定部分包含所有的协变量,与式(13.2)具有相同的形式,对于[LRdist]具有固定效应 $b$,对于所有其他协变量具有党派特定的效应 $\mathbf{g}^s$。由于[1987]与[1992]的系数表示选举特定的常量,因而不需要常数项 $m^s$。

对于离散选择以及排序,我们首先估计不包含潜变量的常规对数概率模型 $M0$,其估计值如表 13.1 所示。我们可以看到,选举及投票者特定的协变量所估计的效应与之前的关于不列颠选举的研究相符合。男性以及年龄大的人投票给保守党的概率高,而上次选举以来对通胀感知度得分高则会损害当权党(保守党)的投票。父亲是手工业者的投票者投给劳工党的概率很高,从这可以显示出社会阶级的影响。在我们这个选举数据中,投票者及候选人特定的协变量[RLdist]的估计值也很有意义:投票者与政党的政治距离越远,投给该政党的可能性越低。

**表 13.1 常规对数概率模型 $M0$ 估计值**

| | | 排 序 | | | | 离散选择 | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Lab 对 Con | | Lib 对 Con | | Lab 对 Con | | Lib 对 Con | |
| | | Est | (SE) | Est | (SE) | Est | (SE) | Est | (SE) |
| $g^s_1$ | [1987] | 0.38 | (0.20) | 0.12 | (0.17) | 0.51 | (0.23) | 0.13 | (0.22) |
| $g^s_2$ | [1992] | 0.51 | (0.20) | 0.13 | (0.18) | 0.63 | (0.24) | -0.13 | (0.23) |
| $g^s_3$ | [Male] | -0.79 | (0.11) | -0.53 | (0.09) | -0.77 | (0.13) | -0.67 | (0.12) |
| $g^s_4$ | [Age] | -0.37 | (0.04) | -0.18 | (0.03) | -0.34 | (0.04) | -0.20 | (0.04) |
| $g^s_5$ | [Manual] | 0.65 | (0.11) | -0.05 | (0.10) | 0.69 | (0.13) | -0.10 | (0.12) |
| $g^s_6$ | [Inflation] | 0.87 | (0.09) | 0.18 | (0.03) | 0.76 | (0.10) | 0.57 | (0.09) |
| $b$ | [LRdist] | -0.62(0.02) | | | | -0.54(0.02) | | | |
| 对数似然值 | | -2 963.68 | | | | -1 957.91 | | | |

资料来源:Skrondal and Rabe-Hesketh (2003b)。

对于随机部分,我们考虑 3 种模型:

(a)政治距离[LRdist]的随机系数模型,引入依赖性,允许 $x^s_{ijk}$ 的效应在投票上变化:$x^s_{ijk}\xi^{(1)}_{ijk}$;在投票者上变化:$x^s_{ijk}\xi^{(2)}_{jk}$;在选区上变化:$x^s_{ijk}\xi^{(3)}_{k}$。在这里,$z^s_{ijk}=x^3_{ijk}$。

(b)单因子模型,引入选举内的依赖性:$\lambda^{s(1)}\eta^{(1)}_{ijk}$,投票者内的依赖性:$\lambda^{s(2)}\eta^{(2)}_{jk}$,以及选区内依赖性:$\lambda^{s(3)}\eta^{(3)}_{k}$。在投票层次上,该模型为公共因子模型,因为 $\epsilon^s_{ijk}$ 可被解释在该层次上的唯一因子。然而,在更高的层次上,则为不存在唯一因子的因子模型。

(c)相关选项特定随机截距模型，引入投票者内的依赖性：$\gamma_{jk}^{s(2)}$，以及选区内依赖性：$\gamma_{k}^{s(3)}$。在投票层次上，我们先不考虑该模型，因为在模型辨识上该模型非常脆弱。

在给定的层次上，如投票者层次，带随机项的(b)模型 $\lambda^{s(2)}\eta_{jk}^{(2)}$，$s=2,3$，嵌套于随机项为 $\gamma_{jk}^{s(2)}$，$s=2,3$ 的随机系数模型(c)中，因为这两个模型中随机项的方差无约束，而协方差在因子模型中被固定为1，但在随机系数模型中无约束。当在多个层次上考虑依赖性时，为了限制所需考虑的模型集合，我们在不同层次上使用了相同类型的模型(如(a)、(b)或(c))。请注意，这是出于实际的考虑；在同一个层次上可以使用不同组合的模型，在不同层次上也可以使用不同类型的模型。甚至，在层次间可以采用无任何参数约束条件的模型。对于多层次模型，采用数字来区分不同的层次，并用括号加字母来区分模型类型，如 $M12$(b)表示投票及投票者层次上的单因子模型。

对于排序问题，所拟合的模型的顺序、参数个数(#Par)及对数似然值如表 13.2 所示(采用 10 点自适应积分法)。

**表 13.2 排序的模型估计(固定部分包含全部协变量)**

| | 随机部分 | | | #Par | 对数似然值 |
|---|---|---|---|---|---|
| | 选 举 | 投票者 | 选 区 | | |
| $M0$ | | | | 13 | -2 963.68 |
| $M1$(a) | $z_{ijk}^{s}\beta_{ijk}^{(1)}$ | | | 14 | -2 945.83 |
| $M1$(b) | $\lambda^{s(1)}\eta_{ijk}^{(1)}$ | | | 15 | -2 842.73 |
| $M2$(a) | | $z_{ijk}^{s}\beta_{jk}^{(2)}$ | | 14 | -2 893.19 |
| $M2$(b) | | $\lambda^{s(2)}\eta_{jk}^{(2)}$ | | 15 | -2 693.78 |
| $M2$(c) | | $\gamma_{jk}^{s(2)}$ | | 16 | -2 645.99 |
| $M3$(a) | | | $z_{ijk}^{s}\beta_{k}^{(3)}$ | 14 | -2 948.44 |
| $M3$(b) | | | $\lambda^{s(3)}\eta_{k}^{(3)}$ | 15 | -2 846.26 |
| $M3$(c) | | | $\gamma_{k}^{s(3)}$ | 16 | -2 844.41 |
| $M12$(a) | $z_{ijk}^{s}\beta_{ijk}^{(1)}$ | $z_{ijk}^{s}\beta_{jk}^{(2)}$ | | 15 | -2 893.19 |
| $M12$(b) | $\lambda^{s(1)}\eta_{ijk}^{(1)}$ | $\lambda^{s(2)}\eta_{jk}^{(2)}$ | | 17 | -2 691.97 |
| $M23$(a) | | $z_{ijk}^{s}\beta_{jk}^{(2)}$ | $z_{ijk}^{s}\beta_{k}^{(3)}$ | 15 | -2 892.87 |
| $M23$(b) | | $\lambda^{s(2)}\eta_{jk}^{(2)}$ | $\lambda^{s(3)}\eta_{k}^{(3)}$ | 17 | -2 630.12 |
| $M23$(c) | | $\gamma_{jk}^{s(2)}$ | $\gamma_{k}^{s(3)}$ | 19 | -2 601.33 |

注：其中，$x_{ijk}^{s}=z_{ijk}^{s}$为左-右翼政治维度的距离。

资料来源：Skrondal and Rabe-Hesketh (2003b)。

我们首先在 $M1$(a)与 $M1$(b)的投票层次中引入潜变量，从表 13.2 中可知，与常规模型 $M0$ 相比，拟合度得到了相当大的提高，说明在给定投票层次中(给定固定部分的协变量)的效用之间存在横截面依赖性(cross-sectional dependence)。

随后，仅在模型 $M2$(a)、$M2$(b)及 $M2$(c)的投票者层次上引入潜变量。与常规模型相比，拟合度也有了很大提高，说明在投票者层次上存在未观测到的异质性，该异质性在投票者内带来纵贯性依赖。我们注意到在过去的政治学研究中，类似于 $M2$(a)的模型非常受关注(如 Rivers, 1988)。

接下来，在 *M*3(a)、*M*3(b)及 *M*3(c)模型中，我们仅在选区层次上加入潜变量，模型的拟合程度再次提高。重要的是，在给定层次上的潜变量不仅在该层次上引入了依赖性，而且在所有更低的层次上引入了依赖性。因而，既然在投票者层次上引入了潜变量，那么在投票层次上的潜变量则显得多余。

为了解决这个问题，我们在模型 *M*12(a)与 *M*12(b)中的投票及投票者层次上都包含潜变量。在投票层次上增加潜变量所带来的拟合度提高可以忽略不计，这说明投票内的横断面依赖性很大程度上是由于投票者层次上的异质性。因而，在接下来的分析中只要投票者层次上包含了潜变量，我们就没有必要在投票层次上再包括潜变量。如果在选区层次上包括潜变量会使得在投票者层次上没有必要包含潜变量的话，那么将会令人惊讶，因为这意味着在给定选区层次效应的条件下，在两次投票中，存在投票者效用间的条件依赖性。因此，正如我们所预期的，模型 *M*23 拟合得比模型 *M*3 更好，这证明在投票者及选区层次都应包含潜变量。至于在 *M*23 的几种模型的选择上，我们看到随机系数模型 *M*23(a)的拟合度比因子模型 *M*23(b)以及随机截距模型 *M*23(c)都差，而在后两个嵌套模型的选择上，更应选择相关随机截距模型。

对于我们最终保留的模型 *M*23(c)的估计值如表 13.3 所示，左半部分为排序，右半部分为离散选择。固定部分的回归系数估计值大于显示在表 13.1 中的常规模型的估计值，这也符合我们的预期。在投票者层次上的随机效应方差大于选区层次上的随机效应方差，这也符合选区内投票者之间的残差变差性高于选区间的投票者之间的残差变差性的预期。

**表 13.3　选项相关的随机截距模型在投票者及选区层次上的估计值 *M*23(c)**

| | 排　序 | | | | 离散选择 | | | |
|---|---|---|---|---|---|---|---|---|
| | Lab 对 Con | | Lib 对 Con | | Lab 对 Con | | Lab 对 Con | |
| | Est | (SE) | Est | (SE) | Est | (SE) | Est | (SE) |
| 固定部分： | | | | | | | | |
| $g_1^s$[1987] | 0.77 | (0.56) | 0.75 | (0.37) | 0.95 | (0.52) | 0.13 | (0.52) |
| $g_2^s$[1992] | 1.28 | (0.59) | 0.78 | (0.39) | 1.32 | (0.54) | 0.30 | (0.55) |
| $g_3^s$[Male] | −0.99 | (0.31) | −0.71 | (0.20) | 1.15 | (0.28) | 0.96 | (0.27) |
| $g_4^s$[Age] | −0.74 | (0.11) | −0.37 | (0.07) | 0.61 | (0.10) | 0.36 | (0.09) |
| $g_5^s$[Manual] | 1.57 | (0.34) | 0.10 | (0.22) | 1.31 | (0.31) | 0.04 | (0.29) |
| $g_6^s$[Inflation] | 1.31 | (0.18) | 0.74 | (0.13) | 1.17 | (0.19) | 0.97 | (0.18) |
| $b$ [LRdist] | −0.79 (0.04) | | | | 0.87 (0.05) | | | |
| 随机部分： | | | | | | | | |
| 投票者层次 | | | | | | | | |
| $\psi^{(2)}_{\gamma_s}$ | 16.13 | (2.05) | 6.03 | (0.90) | 7.43 | (1.62) | 9.11 | (1.61) |
| $\psi^{(2)}_{\gamma^2,\gamma^3}$ | 8.53 (1.15) | | | | 5.90 (1.30) | | | |
| 选区层次 | | | | | | | | |
| $\psi^{(3)}_{\gamma^s}$ | 4.91 | (1.12) | 0.60 | (0.29) | 3.12 | (0.86) | 1.74 | (0.60) |
| $\psi^{(3)}_{\gamma^2,\gamma^3}$ | 1.21 (0.48) | | | | 1.11 (0.60) | | | |
| 对数似然值 | −2 600.90 | | | | 1 748.95 | | | |

资料来源：Skrondal and Rabe-Hesketh (2003b)。

劳工党的随机效应的方差，表示劳工党与保守党效用差异的残差变差性，该值特别大，说明在支持劳工党或保守党上面存在比较强的混杂性（未解释的残差）。劳工党与自由党的随机效应之间的相关说明在控制了协变量之后，倾向于选择劳工党而非保守党的人会同样倾向于选择自由党而非保守党。这也与自由党的位置处于劳工党与保守党之间相符合，同时还表明协变量[LRdist]并不能很好的界定这种顺序。

为了进一步解释该模型随机部分的估计值，我们在表 13.4 中展示该模型所暗含的效用差异的残差（条件依赖于协变量）相关矩阵。为了推导出在给定投票下，某个投票者的效用差异的残相关，效用残差的三维向量模型为

$$\underbrace{\begin{bmatrix} u_{ijk}^{\text{con}} \\ u_{ijk}^{\text{lab}} \\ u_{ijk}^{\text{lib}} \end{bmatrix}}_{\mathbf{u}_{ijk}} - \underbrace{\begin{bmatrix} f_{ijk}^{\text{con}} \\ f_{ijk}^{\text{lab}} \\ f_{ijk}^{\text{lib}} \end{bmatrix}}_{\mathbf{f}_{ijk}} = \underbrace{\begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}}_{\mathbf{Z}} \underbrace{\begin{bmatrix} \gamma_{jk}^{\text{lab}(2)} \\ \gamma_{jk}^{\text{lib}(2)} \end{bmatrix}}_{\gamma_{jk}^{(2)}} + \underbrace{\begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}}_{\mathbf{Z}} \underbrace{\begin{bmatrix} \gamma_{k}^{\text{lab}(3)} \\ \gamma_{k}^{\text{lib}(3)} \end{bmatrix}}_{\gamma_{k}^{(3)}}$$

该效用残差的向量的协方差矩阵就为

$$\text{Cov}(\mathbf{u}_{ijk} - \mathbf{f}_{ijk}) = \mathbf{Z}\boldsymbol{\psi}^{(2)}\mathbf{Z}' + \mathbf{Z}\boldsymbol{\psi}^{(3)}\mathbf{Z}' + \mathbf{I}_3\pi^2/3$$

由于效用残差的向量为

$$\begin{bmatrix} (u_{ijk}^{\text{lab}} - f_{ijk}^{\text{lab}}) - (u_{ijk}^{\text{con}} - f_{ijk}^{\text{con}}) \\ (u_{ijk}^{\text{lib}} - f_{ijk}^{\text{lib}}) - (u_{ijk}^{\text{con}} - f_{ijk}^{\text{con}}) \\ (u_{ijk}^{\text{lab}} - f_{ijk}^{\text{lab}}) - (u_{ijk}^{\text{lib}} - f_{ijk}^{\text{lib}}) \end{bmatrix}$$

定义下述比较矩阵

$$\mathbf{H} = \begin{bmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & -1 \end{bmatrix}$$

效用残差差异的协方差矩阵就为

$$\text{Cov}(\mathbf{H}(\mathbf{u}_{ijk} - \mathbf{f}_{ijk})) = \mathbf{H}\text{Cov}(\mathbf{u}_{ijk} - \mathbf{f}_{ijk})\mathbf{H}'$$

在控制了协变量的情况下，如果自由党的位置处于保守党与劳工党之间，该效用差异的符号为我们所期望的正相关。例如，倾向于劳工党而非保守党（正的 $u_{ijk}^{\text{lab}} - u_{ijk}^{\text{con}}$）的那些人同样会倾向于自由党而非保守党（正的 $u_{ijk}^{\text{lib}} - u_{ijk}^{\text{con}}$）。

**表 13.4　*M*23(c)所表示的效用差异的选区内残相关矩阵(忽略下标 *k*)**

概况

| | | 投票者 $j$ | | 投票者 $j'$ | |
|---|---|---|---|---|---|
| | | 1987 | 1992 | 1987 | 1992 |
| 投票者 $j$ | 1987 | A | | | |
| | 1992 | B | A | | |
| 投票者 $j'$ | 1987 | C | C | A | |
| | 1992 | C | C | B | A |

A：投票内及投票者内

| | $u_{ij}^{\text{lab}} - u_{ij}^{\text{con}}$ | $u_{ij}^{\text{lib}} - u_{ij}^{\text{con}}$ | $u_{ij}^{\text{lab}} - u_{ij}^{\text{lib}}$ |
|---|---|---|---|
| $u_{ij}^{\text{lab}} - u_{ij}^{\text{con}}$ | 1 | | |
| $u_{ij}^{\text{lib}} - u_{ij}^{\text{con}}$ | 0.73 | 1 | |
| $u_{ij}^{\text{lab}} - u_{ij}^{\text{lib}}$ | 0.78 | 0.14 | 1 |

B：投票者内，投票间

| | $u_{i'j}^{\text{lab}} - u_{i'j}^{\text{con}}$ | $u_{i'j}^{\text{lib}} - u_{i'j}^{\text{con}}$ | $u_{i'j}^{\text{lab}} - u_{i'j}^{\text{lib}}$ |
|---|---|---|---|
| $u_{ij}^{\text{lab}} - u_{ij}^{\text{con}}$ | 0.86 | | |
| $u_{ij}^{\text{lib}} - u_{ij}^{\text{con}}$ | 0.63 | 0.67 | |
| $u_{ij}^{\text{lab}} - u_{ij}^{\text{lib}}$ | 0.68 | 0.29 | 0.71 |

C：投票者间（投票内或投票间）

| | $u_{i'j'}^{\text{lab}} - u_{i'j'}^{\text{con}}$ | $u_{i'j'}^{\text{lib}} - u_{i'j'}^{\text{con}}$ | $u_{i'j'}^{\text{lab}} - u_{i'j'}^{\text{lib}}$ |
|---|---|---|---|
| $u_{ij}^{\text{lab}} - u_{ij}^{\text{con}}$ | 0.21 | | |
| $u_{ij}^{\text{lib}} - u_{ij}^{\text{con}}$ | 0.09 | 0.08 | |
| $u_{ij}^{\text{lab}} - u_{ij}^{\text{lib}}$ | 0.22 | 0.06 | 0.27 |

资料来源：Skrondal and Rabe-Hesketh (2003b)。

正如我们所期望的，在给定投票(A)，所暗含的横断面相关大于固定效应模型中所暗含的横断面相关(第一列中的0.5与第二列中的-0.5)。投票者内不同投票之间的纵贯性相关(B)都为正数(固定效应模型意味着零相关)，劳工党与保守党之间效用的差异在投票间相关度最高。同样与我们的期望相符合，同一选区(C)内的投票者之间的相关性大大低于同一个投票者在不同投票之间的相关性。关于自由党与保守党之间差异的相关性比其他都低，这说明在控制了协变量后，这两个党派相对于另外一对党派更难相互区分。

## 13.5 后物质主义：排序的潜类模型

在《宁静革命》中，英格哈特(Inglehart，1977)认为先进工业化社会的政治文化正在发生变革。经济的物质目标以及国家安全正在从人们的基本价值需求中消失。后物质主义的价值正在觉醒，这种价值强调保证自由言论，强调在政治决策中给予人们更多发出声音的机会。随着这种价值观上的变化，这种理论认为，将带来所谓宁静政治，带来政治分歧的变化以及政治参与的变化。英格哈特认为通过要求受访者对物质主义及后物质主义价值观进行排序(而非通过评分法)可以得到最好的量度。在1974/1975年进行的8国调查中(Wieken-Mayser et al.，1979；另见Barnes et al.，1979)——在10.3节中分析过该数据，受访者对下述4种政治价值观进行了排序：

(1)[Order] 保证国家秩序。

(2)[Say] 在政府决策上给予民众更多意见。

(3)[Prices] 抑制通过膨胀。

(4)[Freedom] 保证言论自由。

物质主义者将偏好于[Order]与[Prices]，而后物质主义者将偏好[Say]与[Freedom]。根据克洛恩(Croon，1989)，我们分析如表13.5所示的2 262位德国受访者的数据[1,2]。选项根据上述顺序编号，从而1234表示[Order]、[Say]、[Prices]、[Freedom]的排序顺序。

价值取向异质性的建模可通过假设受访者对政治价值观的“效用”围绕总体均值随机变动。克洛恩(Croon，1989)对于该排序问题介绍了一个探索性潜类分析。价值观$s$与潜类$c$的线性估计量可参数化为

$$\nu_{jc}^{s} = e_{c}^{s}, \quad s = 1,2,3,4, \quad c = 1,\cdots,C \tag{13.5}$$

其中

$$e_{c}^{4} = 0$$

在此，我们估计所有潜类的原点，而非设置其均值为零。单分类、双分类和三分类潜类模型的参数估计值如表13.6所示。

很显然，对于单分类模型，总体上说偏好物质主义价值观，因为对[Order]与[Prices]所估计的对数比率$\hat{e}^1$与$\hat{e}^3$大于[Say]与[Freedom]的$\hat{e}^2$与$\hat{e}^4$(=0)。双分类模型表明占总体大约21%的人是后物质主义者(类别2)，其对价值观1与3的对数比率为负数；占

---

1 本节所分析的数据由巴恩斯(S. H. Barnes)、英格哈特(R. Inglehart)、詹宁斯(M. K. Jennings)以及法拉(B. Farah)处理，来自挪威社会科学数据中心(NSD)。原始数据收集者及挪威社会科学数据中心均不对本分析负责。

2 表13.5的数据可从gllamm. org/books下载。

总体大约79%的人是物质主义者(类别1)。在三类模型中,物质主义者分成了占32%的更加重视[Order](类别3),与占45%的更加重视[Prices](类别1)。后物质主义者所占比例估计为23%。

表13.5 中将三分类模型所估计的后验概率用粗体显示。当[Prices](3)得到很高的排序值时,类别1的后验概率很高;而当[Say](2)与 [Freedom](4)排序值高时,类别2有较高的后验概率;当[Order](1)排序值高时,则为类别3。克洛恩的3分类估计值显示在他的表3中,他使用了不同的参数化方法,其约束为 $\sum_s e_c^s = 0$,而非限制 $e_c^4 = 0$。对于最大的类别,其概率为0.45,得到原点为0.60, -1.07, 1.71 及 -1.24,差不多等同于克洛恩的0.59, -1.07,1.73, -1.25。

**表13.5 物质主义数据及后验概率**

| 数据 | | | | | | 三分类模型结果 | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | 后验概率/% | | | |
| | | | | | | 类别: | 1 | 2 | 3 |
| 排序 | | | | 频次 | 预测频次 | 先验: | 0.45 | 0.23 | 0.32 |
| 1 | 2 | 3 | 4 | 137 | 126 | | 10 | 10 | **80** |
| 1 | 2 | 4 | 3 | 29 | 46 | | 1 | 31 | **67** |
| 1 | 3 | 2 | 4 | 309 | 315 | | 36 | 3 | **61** |
| 1 | 3 | 4 | 2 | 255 | 257 | | 37 | 2 | **61** |
| 1 | 4 | 2 | 3 | 52 | 40 | | 2 | 26 | **73** |
| 1 | 4 | 3 | 2 | 93 | 93 | | 11 | 6 | **83** |
| 2 | 1 | 3 | 4 | 48 | 50 | | 20 | **40** | **40** |
| 2 | 1 | 4 | 3 | 23 | 29 | | 2 | **77** | 22 |
| 2 | 3 | 1 | 4 | 61 | 57 | | **48** | 46 | 6 |
| 2 | 3 | 4 | 1 | 55 | 61 | | 7 | **93** | 0 |
| 2 | 4 | 1 | 3 | 33 | 32 | | 1 | **96** | 3 |
| 2 | 4 | 3 | 1 | 59 | 61 | | 2 | **98** | 0 |
| 3 | 1 | 2 | 4 | 330 | 339 | | **85** | 3 | 12 |
| 3 | 1 | 4 | 2 | 294 | 281 | | **86** | 2 | 12 |
| 3 | 2 | 1 | 4 | 117 | 109 | | **79** | 18 | 4 |
| 3 | 2 | 4 | 1 | 69 | 56 | | 25 | **75** | 0 |
| 3 | 4 | 1 | 2 | 70 | 81 | | **87** | 9 | 4 |
| 3 | 4 | 2 | 1 | 34 | 41 | | 32 | **67** | 0 |
| 4 | 1 | 2 | 3 | 21 | 18 | | 3 | **66** | 32 |
| 4 | 1 | 3 | 2 | 30 | 30 | | 27 | 21 | **51** |
| 4 | 2 | 1 | 3 | 29 | 25 | | 2 | **94** | 4 |
| 4 | 2 | 3 | 1 | 52 | 47 | | 3 | **97** | 0 |
| 4 | 3 | 1 | 2 | 35 | 33 | | **68** | 23 | 9 |
| 4 | 3 | 2 | 1 | 27 | 33 | | 13 | **87** | 0 |

资料来源:Croon (1989)。

克洛恩(Croon, 1989)对这几种不同分类数目的潜类模型进行了评估,他所使用的方法是比较受约束模型与饱和(完整)模型的偏差,而偏差的定义为 2 乘上对数似然值。饱和模型的对数似然值可从原始数据计算得

$$\sum n_R \ln p_R = -6\ 269.52$$

其中,$n_R$ 与 $p_R$ 分别为排序 $R$ 的绝对与相对观测频次,对全部 24 个排序组合进行加总。偏差如表 13.6 所示。

我们将进一步扩展克洛恩的分析,允许潜类概率依赖于下述协变量:

- [Female] 女性,虚拟变量。
- [Age] 年龄分组:15 ~30 (参照组), 31 ~45, 46 ~60 及 60 以上。
- [Education] 受教育程度,定序变量:基础教育程度(参照组)、中等教育程度与大学教育程度。

**表 13.6 克洛恩潜类模型的参数估计值**

| | 单分类 | 双分类 | 三分类 |
|---|---|---|---|
| 类 别 1 | | | |
| 概率 $\pi_1$ | 1 | 0.79 | 0.45 |
| 原 点 | | | |
| $e_1^1$[Order] | 1.16 (0.04) | 1.94 (0.09) | 1.84 (0.15) |
| $e_1^2$[Say] | 0.21 (0.04) | 0.21 (0.05) | 0.17 (0.09) |
| $e_1^3$[Prices] | 1.28 (0.04) | 1.87 (0.09) | 2.96 (0.31) |
| $e_1^4$[Freedom] | 0 | 0 | 0 |
| 类 别 2 | | | |
| 概率 $\pi_2$ | — | 0.21 | 0.23 |
| 原 点 | | | |
| $e_2^1$[Order] | — | -0.87 (0.09) | -0.76 (0.26) |
| $e_2^2$[Say] | — | 0.44 (0.12) | 0.56 (0.12) |
| $e_2^3$[Prices] | — | -0.21 (0.16) | -0.09 (0.19) |
| $e_2^4$[Freedom] | — | 0 | 0 |
| 类 别 3 | | | |
| 概率 $\pi_3$ | — | — | 0.32 |
| 原 点 | | | |
| $e_3^1$[Order] | — | — | 3.14 (0.40) |
| $e_3^2$[Say] | — | — | 0.21 (0.10) |
| $e_3^3$[Prices] | — | — | 1.18 (0.16) |
| $e_3^4$[Freedom] | — | — | 0 |
| 对数似然值 | -6 427.05 | -6 311.69 | -6 281.36 |
| Deviance | 315.05 | 84.32 | 23.58 |
| 自由度 | 20 | 16 | 12 |

我们将结构模型(见4.3.2节)设为

$$\pi_{jc} = \frac{\exp(\mathbf{v}_j'\boldsymbol{\varrho}^c)}{1+\exp(\mathbf{v}_j'\boldsymbol{\varrho}^c)}$$

在总共2 262个受访者中,有2 246位的数据包含完整的协变量。对该子样本重新估计不包含协变量的三分类模型所得到的对数似然值为-6 239.58。估计结果如表13.7所示。允许类概率依赖于协变量将使得对数似然值增加到-6 043.82(损失12个自由度),该估计结果如表13.7所示的$M1$列下面。约束年龄及教育在分类上(得分1,2,…)的效应为线性,仅使得对数似然值降低了2.79(6个自由度),因而,最终保留模型$M2$(显示在表13.7的$M2$列中)。

对于包含协变量的模型$M1$与$M2$,对潜类的解释与不包含协变量的$M0$模型的解释一样,除了在响应模型的所估计的参数上有细微不同以外。有意思的是,女性成为后物质主义者(第二类)的概率更低,与[Age]一样;而[Education]则会增加成为后物质主义者(第二类)的概率。[Education]与[Age]都会增加选择[Order](第三类)而非[Prices]的概率,而[Female]几乎没有效应。

对于模型$M0$与$M2$,分类错误的比例(分类错误率)根据7.4节所估计的$\bar{f}_j$如表13.8所示。如果观测到关于排序的响应,那么受访者可被归入到具有最高后验概率$\omega(e_c|\mathbf{y}_j,\mathbf{x}_j;\hat{\boldsymbol{\theta}})$的类别中。模型$M0$与$M2$所估计的对应的分类错误率分别显示在表格的第二与第四行中。相对于模型$M0$,模型$M2$的分类错误率较低(为0.217),因为它包含潜变量信息。

如果未观测到响应,我们必须给予先验概率$\pi_{jc}$进行分类。在模型$M0$中,这就相当于将所有人分到第一类中(模型概率$\pi_{j1}=0.460$),也就是分类错误率$f_j=0.540$。在模型$M2$中,先验概率使用了协变量信息,因而分类错误率缩减到0.466;见表13.8的第一行与第三行。这种仅基于协变量或"伴随"变量的分类方法有时是必要的,如在目标市场选择中,通常仅有少量的关于产品的排序或偏好的"训练集"以及未来潜在客户的协变量信息(如Wedel, 2002b)。

分类精度也可以用误差削减比例(PRE)的概念来表示,在此,当响应未被观测到时参照的是模型$M0$(第一行)。PRE值显示在表13.8的最后一列中。观测到响应的模型$M2$的PRE值为0.63.

**表13.7　三分类模型参数估计**

| | $M0$<br>包含协变量 | $M1$<br>包含协变量(定类) | $M2$<br>包含协变量(连续) |
|---|---|---|---|
| 响应模型 | | | |
| **类别1** | | | |
| $e_1^1$[Order] | 1.83(0.15) | 1.70(0.14) | 1.71(0.14) |
| $e_1^2$[Say] | 0.18(0.08) | 0.23(0.08) | 0.22(0.09) |
| $e_1^3$[Prices] | 2.93(0.30) | 3.11(0.25) | 3.18(0.26) |
| **类别2** | | | |
| $e_2^1$[Order] | -0.80(0.25) | -0.65(0.15) | -0.64(0.15) |
| $e_2^2$[Say] | 0.54(0.12) | 0.36(0.11) | 0.37(0.10) |

续表

| | $M0$ 包含协变量 | $M1$ 包含协变量（定类） | $M2$ 包含协变量（连续） |
|---|---|---|---|
| $e_2^3$[Prices] | −0.12(0.19) | −0.17(0.13) | −0.15(0.13) |
| **类别 3** | | | |
| $e_3^1$[Order] | 3.14(0.40) | 3.18(0.29) | 3.17(0.29) |
| $e_3^2$[Say] | 0.19(0.10) | 0.20(0.09) | 0.20(0.09) |
| $e_3^3$[Prices] | 1.16(0.17) | 1.40(0.13) | 1.42(0.13) |
| 结构模型 | | | |
| **类别 2** | | | |
| $\varrho_0^2$[Cons] | −0.37(0.34) | 0.31(0.27) | 0.41(0.24) |
| $\varrho_1^2$[Female] | — | −0.90(0.19) | −0.88(0.19) |
| $\varrho_2^2$[Age]† | — | — | −0.94(0.11) |
| $\varrho_3^2$15-30 | — | 0 | — |
| $\varrho_4^2$31-45 | — | −0.97(0.23) | — |
| $\varrho_5^2$46-60 | — | −1.57(0.28) | — |
| $\varrho_6^2$ >60 | — | −3.15(0.51) | — |
| $\varrho_7^2$[Education]† | — | — | 1.51(0.19) |
| $\varrho_8^2$ 基础教育 | — | 0 | — |
| $\varrho_9^2$ 中等教育 | — | 1.65(0.23) | — |
| $\varrho_{10}^2$大学教育 | — | 2.64(0.47) | — |
| **类别 3** | | | |
| $\varrho_0^3$[Cons] | −0.73(0.23) | −0.89(0.37) | −0.69(0.29) |
| $\varrho_1^3$[Female] | — | −0.07(0.16) | −0.07(0.16) |
| $\varrho_2^3$[Age]† | — | — | 0.26(0.08) |
| $\varrho_3^3$15-30 | — | 0 | — |
| $\varrho_4^3$31-45 | — | 0.39(0.31) | — |
| $\varrho_5^3$46-60 | — | 0.90(0.31) | — |
| $\varrho_6^3$ >60 | — | 0.82(0.31) | — |
| $\varrho_7^3$[Education]† | — | — | 0.59(0.19) |
| $\varrho_8^3$ 基础教育 | — | 0 | — |
| $\varrho_9^3$ 中等教育 | — | 0.64(0.22) | — |
| $\varrho_{10}^3$大学教育 | — | 0.84(0.57) | — |
| 对数似然值 | −6 239.58 | −6 043.82 | −6 046.79 |

注：†在定序分类上的线性效应。

表 13.8 分类错误率及误差削减比例(PRE)

| 信 息 | 模 型 | $\tilde{f}_j$ | PRE |
|---|---|---|---|
| 无信息 | *M*0 | 0.540 | |
| 仅有响应 | *M*0 | 0.217 | 0.60 |
| 仅有协变量 | *M*2 | 0.466 | 0.14 |
| 响应及协变量 | *M*2 | 0.202 | 0.63 |

## 13.6 咖啡机的消费偏好：结合选择模型(conjoint choice model)

结合分析法是一种市场研究技术，能够提供有价值的关于市场细分、新产品开发以及价格决策的信息(如 Wedel and Kamakura, 2000)。

在一个真实的购买情景中，顾客在做出最终的购买决定之前，会对一定范围内的产品特性及属性进行考量。结合分析法通过评估各种折中，来判断对于顾客来说哪些特性是最有价值的。只要收集到数据，研究者就可以建立起"模拟的选择"来估计不同特性/属性产品的市场占有。这样，在产品或服务真正进入市场之前，研究者就可以预判哪些产品或服务有可能取得成功。

作为例子，我们分析一个咖啡机的结合选择实验数据[1,2]。通过对专家及消费者的深度访谈，选择了下述 5 种属性来定义一个设想中的咖啡机：

- [Brand] 品牌名称：飞利浦(Philips)、博朗(Braun)、万能(Moulinex)。
- [Capacity] 可盛纳杯数：6, 10, 15。
- [Price] 价格(荷兰盾)：39f, 69f, 99f。
- [Thermos] 是否带保温瓶：是、否。
- [Filter] 是否带特殊过滤：是、否。

通过不齐全设计(incomplete design)，从这 5 个属性的不同程度共得到 16 组搭配(排除不现实的搭配，如都具有上述属性但仅售 39f)。在选择实验中，每 3 个搭配构成一组，每组都有相同的基本属性，受访者被要求从各个组合中进行选择。

相比传统的基于评分的结合分析法，结合选择实验法有不少优点。其中一个优点在于，它更类似真实的购买情景，其次，避免了对评分的个体解释差异问题。

为了建立选择组，搭配用两种不同的方式分成每组两个选项，共 8 组。再往各组加入一个基本选项，得到两大类 8 个选择组，其中每个选择组含 3 个选项，其分组结果如表 13.9 所示。在荷兰的一个大型购物商场征集了 185 位受访者。这些受访者被随机的分成两组(分别为 94 人与 91 人)，每组回答上述 8 个选择组的其中一类。

对于第 $j$ 个受访者，第 $s$ 个选择(或搭配)的属性 $\mathbf{x}^s$ 构成向量，仅包含该向量的固定效应的多项式对数概率模型的线性估计量为

$$\nu_j^s = \mathbf{x}^{s\prime}\mathbf{b}$$

---

1 我们感谢迈克·维德尔(Michel Wedel)提供该数据，并随 GLIMMIX 程序一起发布 (Wedel, 2002b)。

2 该数据可从 gllamm.org/books 下载。

表 13.9 结合选择实验的候选集

| | 选项 1 | | | | | 选项 2 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 组别 | Brand | Cap. | Pr. | Th. | Fi. | Brand | Cap. | Pr. | Th. | Fi. |
| | | | | | 第一组搭配 | | | | | |
| 1 | Philips | 10 | 69 | — | Fi | Braun | 15 | 69 | Th | Fi |
| 2 | Braun | 6 | 69 | — | — | Moulinex | 10 | 69 | Th | Fi |
| 3 | Braun | 10 | 39 | — | — | Braun | 10 | 99 | Th | Fi |
| 4 | Philips | 6 | 39 | Th | Fi | Braun | 10 | 39 | Th | — |
| 5 | Philips | 10 | 69 | Th | — | Moulinex | 15 | 39 | — | Fi |
| 6 | Braun | 6 | 69 | — | Fi | Moulinex | 10 | 69 | — | — |
| 7 | Philips | 15 | 99 | — | — | Moulinex | 10 | 99 | Th | — |
| 8 | Braunn | 15 | 69 | Th | — | Braun | 6 | 99 | — | Fi |
| | | | | | 第二组候搭配 | | | | | |
| 1 | Philips | 10 | 69 | Th | — | Moulinex | 10 | 69 | Th | Fi |
| 2 | Philips | 15 | 99 | — | — | Braun | 15 | 69 | Th | Fi |
| 3 | Braun | 10 | 39 | — | — | Moulinex | 15 | 39 | — | Fi |
| 4 | Braun | 15 | 69 | Th | — | Braun | 10 | 99 | — | Fi |
| 5 | Philips | 10 | 69 | — | Fi | Moulinex | 6 | 99 | Th | — |
| 6 | Braun | 6 | 69 | — | — | Braun | 10 | 99 | Th | Fi |
| 7 | Braun | 6 | 69 | — | — | Moulinex | 10 | 69 | — | — |
| 8 | Philips | 6 | 39 | Th | Fi | Braun | 10 | 39 | Th | — |

| | 选项 3(基　准) | | | | |
|---|---|---|---|---|---|
| 组别 | Brand | Cap. | Pr. | Th. | Fi. |
| 1 | Philips | 6 | 69 | — | Th |

请注意,为了模型的辨识,忽略了截距 $m^s$,因为协变量 $\mathbf{x}^s$ 并未在在受访者上变化。去除常数项还有另外一个优点,即我们可以用来估计没有包括在数据中的其他搭配。这个模型实际上是不存在的,因为它假设所有的受访者在咖啡机上都有相同的效用均值。更真实的模型应当允许不同受访者在属性的系数上不同,以反映不同的偏好或“口味”。因而,线性估计量则为

$$\nu_j^s = \mathbf{x}^{s\prime}\mathbf{b} + \mathbf{x}^{s\prime}\gamma_j$$

如果相信市场是由具有不同偏好的“细分市场”所构成的,那么可用离散随机系数向量设定一个潜类模型

$$\gamma_j = \mathbf{e}_c, \quad c = 1, \cdots, C$$

其中,概率为 $\pi_c$,$c$ 表示不同的细分市场。该模型也经常被称为混合回归模型(mixture regression model)。我们将从模型中移除固定部分 $\mathbf{x}^s$,而非约束 $\sum_c \pi_c \mathbf{e}_c = \mathbf{0}$,这使得对分类 $c$ 的线性估计量为

$$\nu_{jc}^s = \mathbf{x}^{s\prime}\mathbf{e}_c$$

或者,可以假设随机偏好 $\gamma_j$ 为连续变量。哈埃耶等(Haaijer et al., 1998)提出了一个随机系数多项式正态概率模型,其中,$\gamma_j$ 是多元正态分布,并且其协方差结构为

$$\text{Cov}(\gamma_j) = \mathbf{QQ}'$$

其中,$\mathbf{Q}$ 为 $8 \times T$ 的矩阵,并且为了模型辨识,对 $Q$ 必须做一些约束。然而,即使做了约束,他们发现当 $T > 1$ 时,仍然很难辨识模型。当 $T = 1$ 时,该模型的多项式对数概率模型版本可用 GRC 方程式表示为

$$\nu_j^s = \mathbf{x}^{s\prime}\mathbf{b} + \eta_j \mathbf{x}^{s\prime}\boldsymbol{\lambda}, \quad \text{Var}(\eta_j) = 1$$

其中,我们设置 $\gamma_j = \eta_j \boldsymbol{\lambda}$,将维度从 8 降低到 1,大大简化了估计。

对于单分类模型、双分类模型及随机系数模型的估计结果显示在表 13.10 中。单分类模型的系数说明 Braun 为最不流行的品牌;10 杯容量是最适合的,其次为 15 杯,再次为 6 杯;更偏好便宜的咖啡机(36f 或 69f),而非贵的(99f);更偏好带有过滤及保温瓶的咖啡机。

双分类模型拟合地比单分类模型更好(对数似然值相差 188)。第一分类细分市场的大小估计为 72%,而第二分类为 28%。第一分类细分市场不注重品牌,而第二分类更偏好 Philips,不喜欢 Braun。第一分类比第二分类更偏好 10 杯容量,更关心价格,以及都对保温瓶有更强的偏好。

随机系数模型说明受访者主要在偏好 Philips 而非 Braun,不喜欢 6 杯容量,以及价格敏感性上面,有很强的差异性。值得一提的是,该独立多项式对数概率模型的对数似然值非常接近于哈埃耶等(Haaijer et al., 1998)用模拟最大似然估计法估计的独立多项式正态概率模型(分别为 -1 298.7 与 -1 299.9)。同样,随机系数模型的对数概率版本与正态概率版本的对数似然值也非常接近(分别为 -1 086.1 与 -1 086.6)。

**表 13.10 结合选择分析估计值**

| | 潜类模型 | | | | | 随机系数模型 | | |
|---|---|---|---|---|---|---|---|---|
| | | $M1$ 单分类 | | $M2$ 双分类 | | | $M3$ | |
| 变 量 | 参 数 | Est | (SE) | Est | (SE) | 参 数 | Est | (SE) |
| [Brand] | | | | | | | | |
| Philips | $e_1^1$ | -0.25 | (0.11) | -0.37 | (0.17) | $b_1$ | -0.31 | (0.15) |
| Braun | $e_2^1$ | -0.62 | (0.11) | -0.40 | (0.16) | $b_2$ | -0.65 | (0.15) |
| Moulinex | $e_3^1$ | 0 | | 0 | | $b_3$ | 0 | |
| [Capacity] | | | | | | | | |
| 6 | $e_4^1$ | -1.54 | (0.12) | -2.48 | (0.21) | $b_4$ | -2.08 | (0.20) |
| 10 | $e_5^1$ | 0.03 | (0.10) | 0.06 | (0.14) | $b_5$ | 0.13 | (0.13) |
| 15 | $e_6^1$ | 0 | | 0 | | $b_6$ | 0 | |
| [Price] | | | | | | | | |
| 39 | $e_7^1$ | 1.00 | (0.15) | 1.97 | (0.34) | $b_7$ | 1.52 | (0.27) |
| 69 | $e_8^1$ | 1.06 | (0.11) | 1.48 | (0.17) | $b_8$ | 1.21 | (0.16) |
| 99 | $e_9^1$ | 0 | | 0 | | $b_9$ | 0 | |

续表

| 变量 | 参数 | 潜类模型 M1 单分类 Est | (SE) | 潜类模型 M2 双分类 Est | (SE) | 随机系数模型 参数 | M3 Est | (SE) |
|---|---|---|---|---|---|---|---|---|
| [Thermos] | | | | | | | | |
| 有 | $e_{10}^1$ | 0.62 | (0.09) | 1.14 | (0.18) | $b_{10}$ | 1.09 | (0.15) |
| 无 | $e_{11}^1$ | 0 | | 0 | | $b_{11}$ | 0 | |
| [Filter] | | | | | | | | |
| 有 | $e_{12}^1$ | 0.68 | (0.08) | 0.92 | (0.12) | $b_{12}$ | 0.97 | (0.11) |
| 无 | $e_{13}^1$ | 0 | | 0 | | $b_{13}$ | 0 | |
| $\ln[\pi_1/(1-\pi_1)]$ | $\varrho_0^1$ | 0 | | 0.92 | (0.21) | — | | |
| [Brand] | | | | | | | | |
| Philips | $e_1^2$ | | | 0.12 | (0.21) | $\lambda_1$ | 0.51 | (0.18) |
| Braun | $e_2^2$ | | | -1.43 | (0.31) | $\lambda_2$ | -0.51 | (0.20) |
| Moulinex | $e_3^2$ | | | 0 | | $\lambda_3$ | 0 | |
| [Capacity] | | | | | | | | |
| 6 | $e_4^2$ | | | -0.25 | (0.26) | $\lambda_4$ | 1.72 | (0.22) |
| 10 | $e_5^2$ | | | 0.07 | (0.25) | $\lambda_5$ | 0.19 | (0.21) |
| 15 | $e_6^2$ | | | 0 | | $\lambda_6$ | 0 | |
| [Price] | | | | | | | | |
| 39 | $e_7^2$ | | | -0.49 | (0.32) | $\lambda_7$ | -1.71 | (0.30) |
| 69 | $e_8^2$ | | | -0.04 | (0.22) | $\lambda_8$ | -1.11 | (0.21) |
| 99 | $e_9^2$ | | | 0 | | $\lambda_9$ | 0 | |
| [Thermos] | | | | | | | | |
| 有 | $e_{10}^2$ | | | 0.35 | (0.20) | $\lambda_{10}$ | -0.60 | (0.19) |
| 无 | $e_{11}^2$ | | | 0 | | $\lambda_{11}$ | 0 | |
| [Filter] | | | | | | | | |
| 有 | $e_{12}^2$ | | | 1.00 | (0.20) | $\lambda_{12}$ | 0.08 | (0.14) |
| 无 | $e_{13}^2$ | | | 0 | — | $\lambda_{13}$ | 0 | |
| 对数似然值 | | -1 298.71 | | -1 110.63 | | | -1 086.07 | |

哈埃耶等(Haaijer et al., 1998)还考虑了对产品采用不同介绍方法的选择模拟。他们使用表 13.11 所列的 4 种搭配来生成 3 种"经营相关"的情景:

**表 13.11 市场模拟搭配**

| 搭配 | 品牌 | Cap. | Pr. | Th. | Fi. | 搭配 | 品牌 | Cap. | Pr. | Th. | Fi. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **P1** | Philips | 10 | 39 | — | — | **B2** | Braun | 15 | 69 | Th | — |
| **M3** | Moulinex | 15 | 69 | — | — | **P4** | Philips | 10 | 69 | — | Fi |

• 产品升级:当前市场存在两种产品:Philips(P1) 与 Braun (B2)。Philips 通过增加过滤器及提高价格对其现有产品(从 P1 到 P4)进行了升级。

• 产品线延伸:当前市场存在两种产品,Philips(P1) 与 Braun (B2)。Philips 在其现有产品中又增加了一个新产品(P4)。

• 引入模仿品牌:当前市场存在两种产品,Philips (P1) 与 Braun (B2)。第三个品牌 Moulinex 推出了新产品(M3),并与当前的市场领先者 Braun 的产品类似,但是缺少保温瓶。

对于这 3 种情景,各个模型对市场占有率的估计结果如表 13.12 所示。在单分类模型及双分类模型中,产品升级将促进市场占有率的增加,高于随机系数模型。Philips 的产品线延伸将导致 Braun 市场占有率下降的程度在单分类模型及双分类模型中高于随机系数模型。同样,新品牌的加入造成 Braun 市场占有率下降的程度在单分类模型及双分类模型中也高于随机系数模型。该虚拟的例子演示了通过假设不同的依赖结构对于预测所造成的巨大影响。

**表 13.12 市场占有率估计百分比**

| 产品升级 | | | | 产品线延伸 | | | | 引入模仿品牌 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 产品 | *M1* | *M2* | *M3* | 产品 | *M1* | *M2* | *M3* | 产品 | *M1* | *M2* | *M3* |
| 之前: | | | | | | | | | | | |
| P1 | 41.5 | 45.8 | 43.9 | P1 | 41.5 | 45.8 | 43.9 | P1 | 41.5 | 45.8 | 43.9 |
| B2 | 58.5 | 54.2 | 56.1 | B2 | 58.5 | 54.2 | 56.1 | B2 | 58.5 | 54.2 | 56.1 |
| 之后: | | | | | | | | | | | |
| P4 | 59.8 | 59.4 | 55.2 | P1 | 22.2 | 21.6 | 18.6 | P1 | 26.2 | 30.3 | 30.4 |
| B2 | 40.2 | 40.6 | 44.8 | B2 | 31.3 | 31.4 | 37.1 | B2 | 37.0 | 39.4 | 43.8 |
| | | | | P4 | 46.5 | 47.1 | 44.3 | M3 | 36.8 | 30.3 | 25.7 |

## 13.7 总结与进一步阅读

在本章中,我们基于连续或离散潜变量的对数概率回归模型分析了几种比较响应模型。第一个应用的例子是关于不列颠普选的离散选择及排序的纵贯性数据。模型包含选项及单元特定的协变量以及在不同层次上的潜变量。第二个例子是关于德国人的物质主义价值观与后物质主义价值观。通过包含协变量到潜类模型中扩展了之前对这个问题的分析。第三个例子基于结合选择实验,让受访者在几种咖啡机中进行选择。根据实验设计,咖啡机不同属性的搭配构成了选择集,来帮助市场研究者在产品向市场推出前对产品是否能够成功进行评估。关于当代市场细分的问题,请见 Wedel and Kamakura (2000)。

在本章所讨论的所有例子中,题器的响应都是决策,但实际并不限于此。关于多分类响应的一个例子为眼睛颜色或诊断。杨 (Yang, 2001) 及斯科隆多和拉贝-赫斯基思 (Skrondal 与 Rabe-Hesketh, 2003c) 使用了一个三层次多项式对数概率模型来分析医生诊断决策的质量。成对比较数据在体育比赛中,如象棋或足球比赛,也应用得越来越多;而排序可用于赛马的名次。

除了在本章开头所介绍的参考文献以外,还可参见哈泽尔等(Hartzel et al., 2001)以及特雷恩(Train, 2003)的综述。关于多分类响应及排序的潜变量模型的综述,请见斯科隆多和拉贝-赫斯基思(Skrondal and Rabe-Hesketh, 2003b)。斯科隆多和拉贝-赫斯基思(Skrondal and Rabe-Hesketh, 2003a)给了很多介绍性的分析。高岭(Takane, 1987)及博肯哈特(Böckenholt, 2001a)讨论了关于成对比较的潜变量模型。

# 14 多过程响应与混合响应

## 14.1 介　绍

在前面的应用章节中,我们讨论了特殊类型的响应。在本章,我们将拓展一般模型框架,并讨论来自多过程的响应变量以及混合类型变量的应用。本章所讨论的混合类型变量包括连续变量与二项变量、二项变量与事件数变量、连续变量与连续时间生存变量以及二项变量与连续时间生存变量。重要的是,我们将看到,对于这些问题有时无法简单地通过分解(即对不同过程分别建模)来实现。与其他章节一样,我们将使用带参数分布的连续潜变量,以及可解释为潜类或作为未知分布的"非参"估计量的离散潜变量。另外,还将展示这种用潜变量对其他变量回归的结构模型的优点。

## 14.2 饮食与心脏病:协变量测量误差模型

### 14.2.1 介　绍

根据拉贝-赫斯基思等(Rabe-Hesketh et al., 2003ab)的研究,我们将估计膳食纤维(粗纤维)摄入量对冠心病(CHD)的效应。数据来自 337 位中年男性,从 1956—1966 年间进入调查,并一直跟踪到 1976 年(Morris et al., 1977)[1]。数据包含两种职业类型,银行职员以及伦敦公交系统职员(司机与售票员)。在首次调查时,受访者被问到他在 7 天内所消耗的食物总量,从中可以得到卡路里总量以及脂肪与纤维的含量。在 6 个月后,用同样的方式再次测量了 76 位银行职员的食物消耗量。是否患有冠心病则通过职员记录得到;如果该受访者已退休,则通过直接询问受访者或从注册机构查询得到。

该研究所用的探索性变量如下:

- [Age] 年龄。
- [Transp] 虚拟变量,公交职员。

### 14.2.2 包含协变量测量误差的对数概率模型

我们接下来将使用对数概率模型估计纤维摄入量(接触)与心脏病的关系,并且考虑到该接触测量所存在的非精确性,以及存在重复测量的子样本。这个问题可以通过对未观测到的真实接触引入一个潜变量,并设定 3 个子模型(根据 Clayton (1992)的定义)来

---

1 感谢大卫·克雷顿(David Clayton)提供该数据。

实现:一个接触模型、一个测量模型和一个疾病模型。

***接触模型***

对受访者$j$,其真实的纤维摄入量$\eta_j$用结构模型可表达为

$$\eta_j = \mathbf{x}_j'\gamma + \zeta_j \tag{14.1}$$

其中,协变量$\mathbf{x}_j$为[Age]、[Transp]以及它们的交互项。一般来说,正态接触分布被假设为服从$\zeta_j \sim N(0,\psi)$,但我们在此将使用4.4.2节所介绍的非参接触分布。

***测量模型***

经典测量模型假设对于受访者$j$,其第$i$次纤维测量,$y_{ij}$,对于真实纤维摄入量$\eta_j$的测量误差$\epsilon_{ij}$服从正态分布,即

$$y_{ij} = \eta_j + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0,\theta)$$

其中,$\epsilon_{ij}$独立于$\eta_j$(以及接触模型中的$\zeta_j$)。有些子样本存在两次测量,但只进行了一次测量的受访者的信息我们也照样使用。为了允许纤维测量的"漂移",我们对于第二次测量建立一个虚拟变量[Drift]。

***疾病模型***

疾病模型对于是否存在心脏病$D_j$设定一个关于真实纤维摄入量的对数概率模型,即

$$\text{log it}[\Pr(D_j = 1 \mid \eta_j)] = \mathbf{x}_j'\boldsymbol{\beta} + \lambda\eta_j$$

其中,因子负荷$\lambda$表示真实接触对于是否发病的对数比率效应。在疾病模型和接触模型中都包含同样的协变量$\mathbf{x}_j$,以允许协变量对发病的直接效应和间接效应。我们还分析$\boldsymbol{\beta}=\mathbf{0}$的模型,即协变量对于心脏病仅有通过真实纤维摄入量的间接效应。这两个模型如图14.1所示。

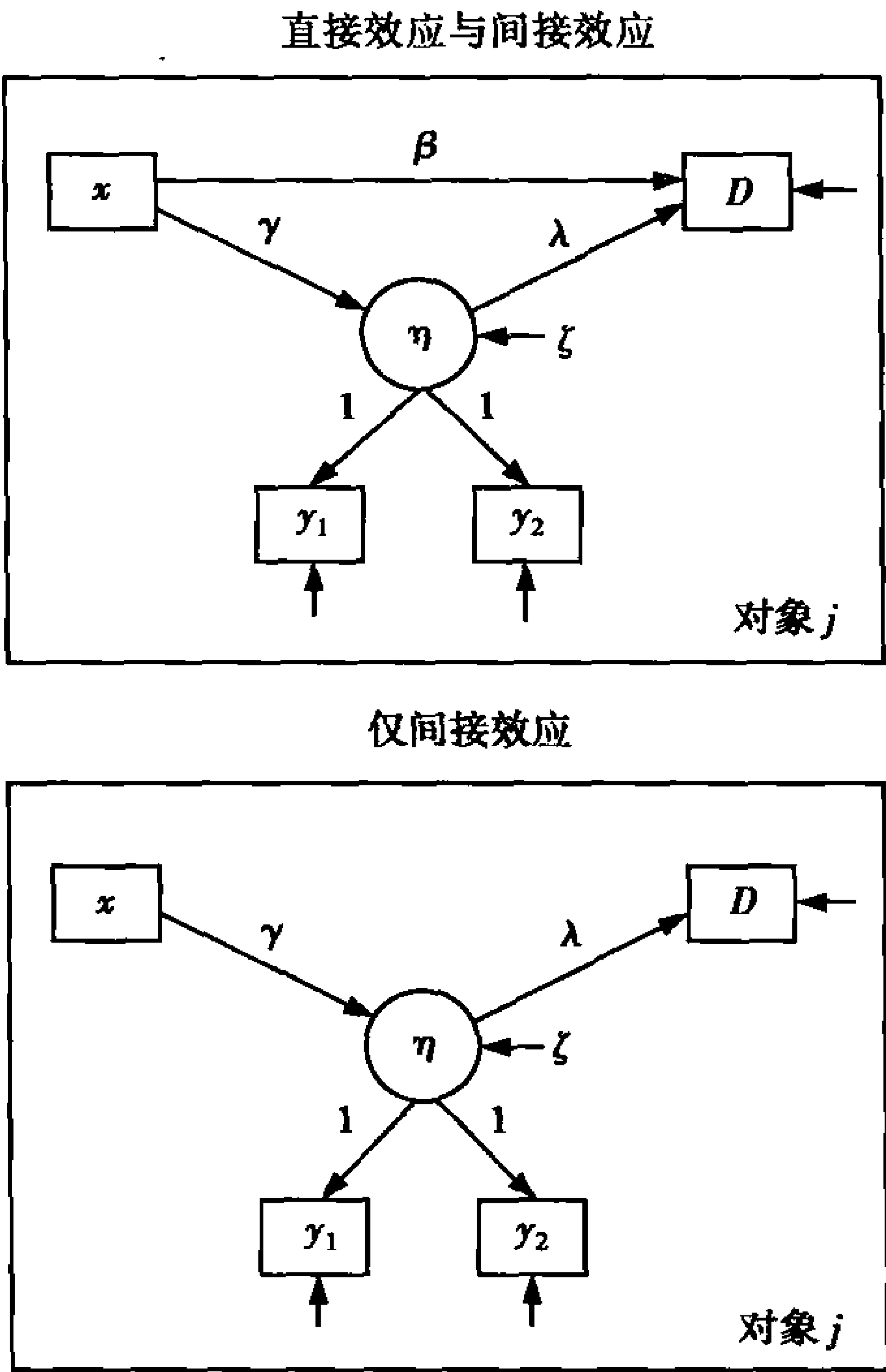



**图14.1 协变量测量误差模型的路径图**

***联合模型***

设$i=1,2$表示纤维测量的次数,$i=3$表示发病响应,另外定义对应的虚拟变量$d_{1i}$,$d_{2i}$及$d_{3i}$。测量与发病的结合响应模型用GRC方程式可写为

$$\begin{aligned}\nu_{ij} &= (d_{1i}+d_{2i})\eta_j + d_{3i}[\mathbf{x}_j'\boldsymbol{\beta}+\lambda\eta_j]\\ &= d_{3i}\mathbf{x}_j'\boldsymbol{\beta}+\eta_j[(d_{1i}+d_{2i})+\lambda d_{3i}]\end{aligned}$$

接触的结构模型仍然是式(14.1)。

请注意,测量误差被假设为**非差异性**的,即在给定真实接触 $\eta_j$(在路径图中,$y_{ij}$ 与 $D_j$ 之间不存在直接路径)的条件下,测量误差条件依赖于发病状态 $D_j$。该结合模型是3.5节所讨论的MIMIC模型的一般化,它允许混合类型的响应变量(饮食测量为连续变量,心脏病为二项变量)以及协变量对响应变量的直接效应。

协变量对心脏病的包含或不包含直接效应,以及基于正态或非参接触分布的参数估计结果显示在表14.1中。这四种模型所估计的真实纤维摄入量对心脏病的效应都非常接近,每天每克纤维摄入量的优比值为0.87。纤维摄入量看起来对于防止心脏病有显著效应。然而,我们并没有控制锻炼这个混淆因素,因为它既能增加食物摄入量(包括纤维),也能降低患心脏病的风险(Morris et al.,1977)。伦敦公交职员比银行职员摄入的纤维量少,纤维摄入量随年龄增加而减少,并且,在职业与年龄之间存在实质性的交互。排除这些变量对心脏病的直接效应,在正态以及非参接触分布假设下,都仅使得对数似然值略有降低。因而,协变量似乎只通过纤维摄入量间接地对患心脏病风险产生效应,正如图14.1所示的下半部分所显示的那样。

对于只包含间接效应的模型,在给定协变量条件下,当假设接触服从正态分布时,纤维测量信度的估计值为0.77,用NPMLE估计时,测量信度估计值为0.80。这个略高的NPMLE估计值与拉贝-赫斯基思等(Rabe-Hesketh et al., 2003a)以及胡等(Hu et al., 1998)所报告的模拟结果一致。

**表14.1　心脏病数据参数估计值**

| | 仅有间接效应 | | | | 直接效应与间接效应 | | | |
|---|---|---|---|---|---|---|---|---|
| | 积　分 | | NPMLE | | 积　分 | | NPMLE | |
| | Est | (SE) | Est | (SE) | Est | (SE) | Est | (SE) |
| 接触模型 | | | | | | | | |
| $\gamma_1$[Transp] | -1.66 | (0.64) | -1.12 | (0.44) | -1.68 | (0.64) | -1.14 | (0.44) |
| $\gamma_2$[Age] | -0.21 | (0.10) | -0.29 | (0.06) | -0.21 | (0.10) | -0.28 | (0.06) |
| $\gamma_3$[Age] ×[Transp] | 0.17 | (0.11) | 0.22 | (0.07) | 0.17 | (0.11) | 0.22 | (0.07) |
| Var($\zeta_i$) | 23.66 | (2.53) | 24.94 | (—) | 23.64 | (2.52) | 24.98 | (—) |
| 测量模型 | | | | | | | | |
| $\alpha_0$[Const] | 17.93 | (0.49) | 17.58 | (0.40) | 17.95 | (0.49) | 17.60 | (0.40) |
| $\alpha_1$[Drift] | 0.24 | (0.42) | 0.16 | (0.38) | 0.23 | (0.42) | 0.15 | (0.38) |
| $\theta$ | 6.95 | (1.14) | 6.13 | (0.85) | 6.95 | (1.14) | 6.13 | (0.85) |
| 疾病模型 | | | | | | | | |
| $\lambda$ | -0.13 | (0.05) | -0.15 | (0.06) | -0.13 | (0.05) | -0.15 | (0.06) |
| $\beta_0$[Const] | -2.08 | (0.21) | -2.07 | (0.21) | -1.92 | (0.28) | -1.90 | (0.28) |
| $\beta_1$[Transp] | — | | — | | -0.26 | (0.34) | -0.27 | (0.34) |
| $\beta_2$[Age] | — | | — | | 0.04 | (0.06) | 0.04 | (0.06) |
| $\beta_3$[Age] ×[Transp] | — | | — | | -0.03 | (0.06) | -0.03 | (0.07) |
| 对数似然值 | -1 373.33 | | -1 320.25 | | -1 372.35 | | -1 319.79 | |

NPMLE 解法需要 6 个概率质量，并显示在图 14.2 上半部分的仅包含间接效应的模型中(包含直接效应的模型的分布与其非常类似)。正如所料，该分布向右偏斜。另外一个证明真实纤维摄入量分布为非正态的证据是，在释放正态假设后，对数似然值急剧增加(NPMLE 解法多了 9 个参数，对数似然值增加 53)。该图下半部分显示的是真实纤维摄入量在正态及非参分布条件下所估计的累计分布。

接触模型的干扰项 $\zeta_j$ 的经验贝叶斯估计如图 14.3 所示，包括 NPMLE 与正态两种情形。在估计值较高的地方，两者的差异增大，NPMLE 估计值较高。拉贝-赫斯基思等(Rabe-Hesketh et al., 2003a)的模拟结果显示，当真实分布高度偏斜时，NPMLE 估计优于正态假设。特别要指出的是，当真实值很大时，参数经验贝叶斯估计值会得到过于缩小的结果。

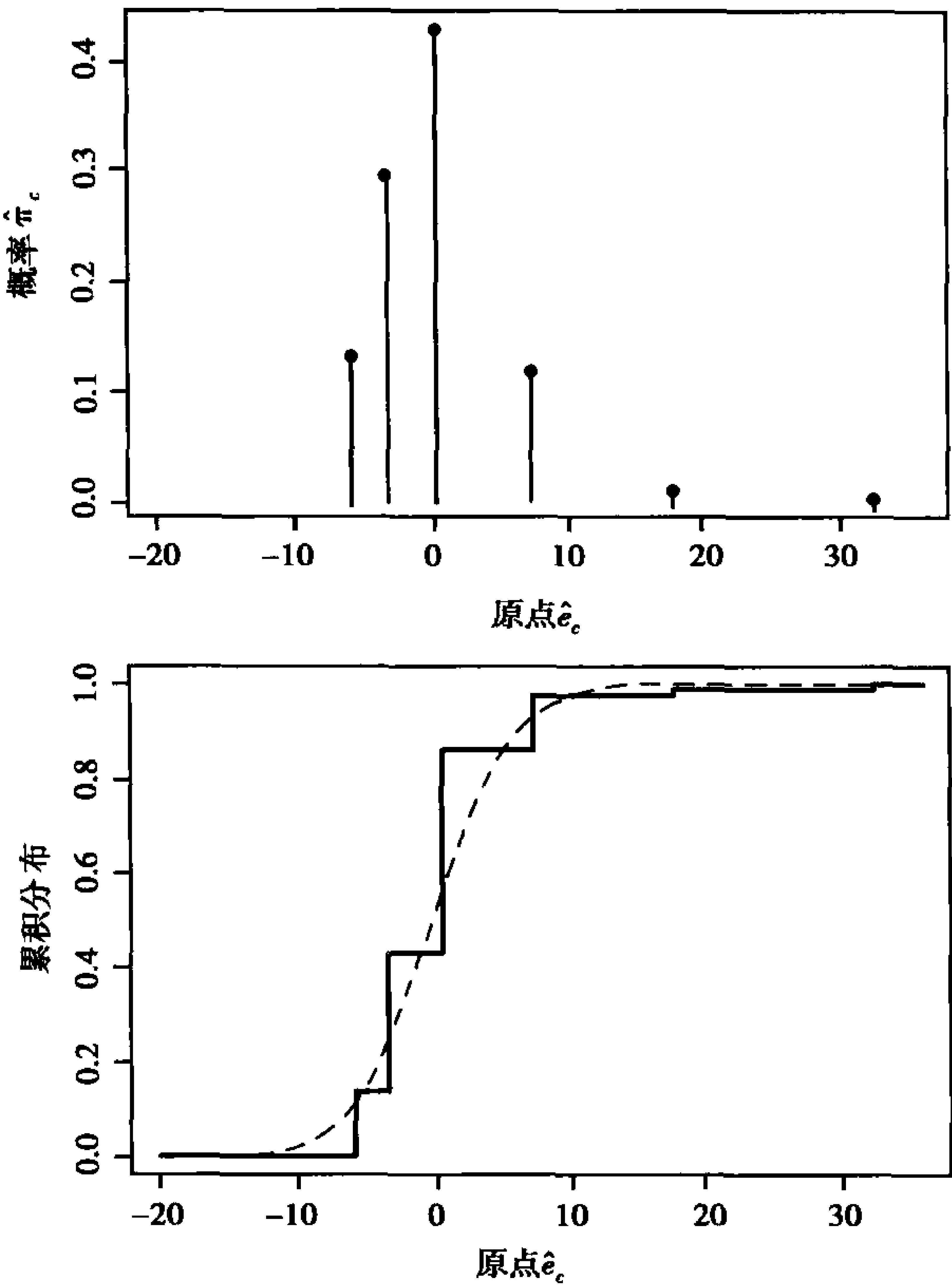



图 14.2 NPMLE 概率质量以及 NPMLE 与正态分布的累计分布

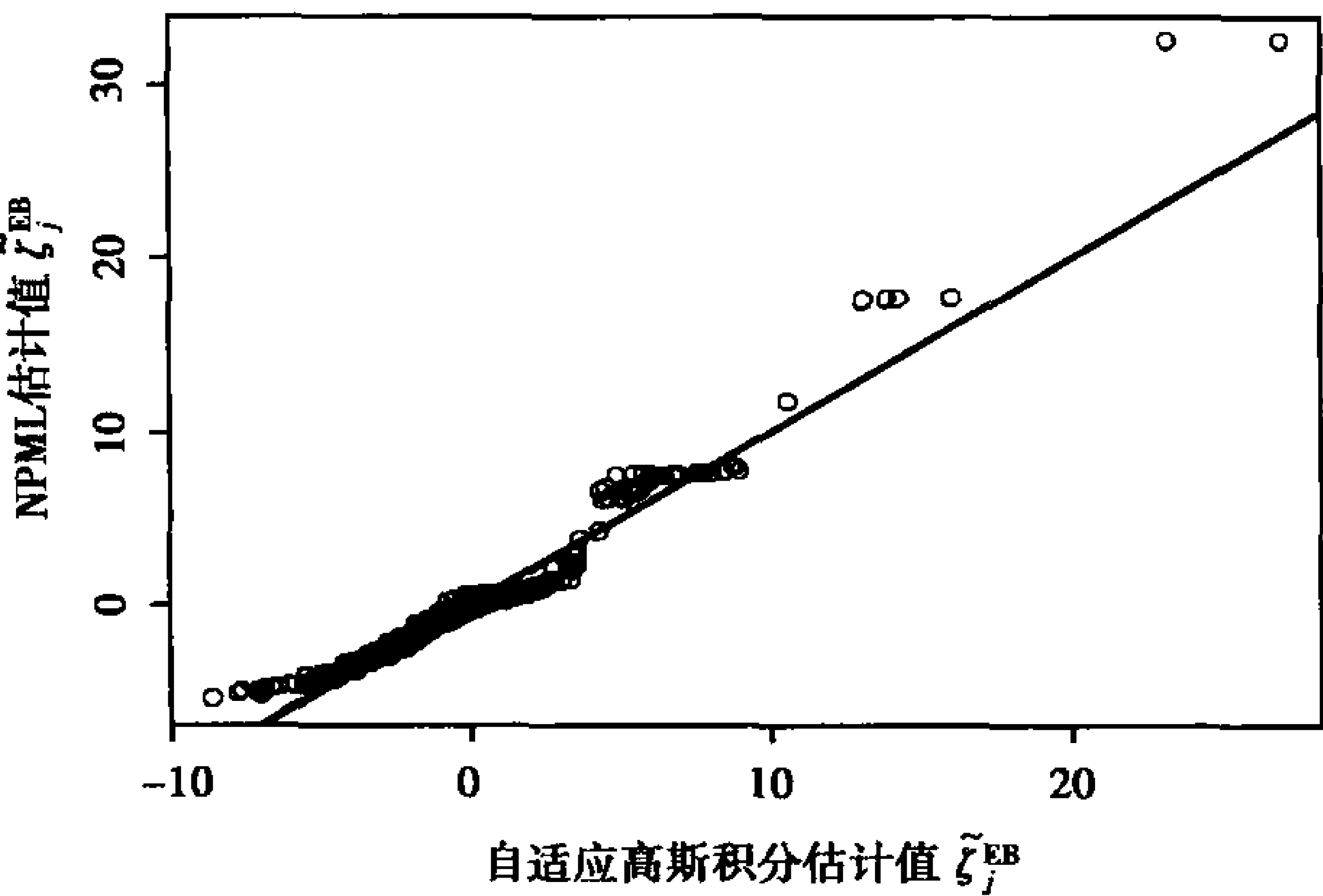



图 14.3 正态和非参接触分布假设下，真实纤维摄入量残差经验贝叶斯估计

## 14.3　生殖器疱疹与子宫颈癌：一项病例对照研究的潜类协变量测量误差模型

### 14.3.1　介　绍

在某些情况下，基于一个二项响应进行分层抽样。这种回溯性设计在流行病学中称为病例对照研究（如 Breslow and Day, 1980; Schlesselman, 1982），在经济学中称为基于选择的抽样（如 Manski, 1981）。抽样通常包含所有的“案例”（在流行病学中为病人），以及一个尽可能多的“控制组”（健康的人）随机样本。其重点在于，当发病率较低时，效力的降低比前瞻性同期群设计（prospective cohort design）相对较小。与前瞻性同期群设计相比，它的一个很重要的优点在于可以得到更精确的协变量信息，因为只需对控制组的样本进行测量。

研究显示，对于这种回溯性设计，可以把它当做前瞻性设计并使用对数概率模型（如 Farewell, 1979），能得到除截距之外的合理估计。重要的是，卡洛尔等（Carroll et al., 1995b）指出在病例对照研究中，带有协变量测量误差的前瞻性分析通常能得到一致的估计量以及近似的正确标准误。特别是，对于定类的接触以及非差异性测量误差，沙特恩和库珀（Satten and Kupper, 1993）以及卡洛尔等（Carroll et al., 1995b）已经说明通过最大化前瞻性似然能够得到正确的推论。在本小节中，我们将采用这个方法，在设定似然的时候忽略回溯性设计。

表 14.2　子宫颈癌病例对照研究数据

| | 真实接触 | 测量到的接触 | |
|---|---|---|---|
| [Case] | [TrueE] | [MeasE] | 事件数 |
| 校验样本 | | | |
| 1 | 0 | 0 | 13 |
| 1 | 0 | 1 | 3 |
| 1 | 1 | 0 | 5 |
| 1 | 1 | 1 | 18 |
| 0 | 0 | 0 | 33 |
| 0 | 0 | 1 | 11 |
| 0 | 1 | 0 | 16 |
| 0 | 1 | 1 | 16 |
| 不齐全样本 | | | |
| 1 | | 0 | 318 |
| 1 | | 1 | 375 |
| 0 | | 0 | 701 |
| 0 | | 1 | 535 |

资料来源：Carroll et al.（1993）。

希尔德斯海姆等(Hildesheim et al., 1991)进行了一项病例对照研究,分析Ⅱ型单纯疱疹病毒与子宫颈癌的可能关联。接触的测量采用一种不太精确的蛋白印迹法(western blot)。为了研究分类错误,采用精炼的**蛋白印迹法**分别从两个组中抽取一个随机子样本,也就是测量的金标准——校验样本。我们将该测量当做“真实接触”[TrueE],而那个不太精确的**蛋白印迹法**的结果当做“测量接触”[MeasE]。该数据[1]由卡洛尔等(Carroll et al., 1993)及其他一些研究者分析过,如表14.2所示。

当然,我们可以基于校验样本对真实接触的优比进行有效估计。然而,这种方法效率不高,因为包含在大量不齐全样本中的信息被忽略掉了。

### 14.3.2 潜类对数概率模型

我们可以基于全部信息,来估计真实接触的优比,方法是像前一节那样设定3个构成模型:一个接触模型、一个测量模型和一个疾病模型。

***接触模型***

设 $X_{1j}$ 表示校验样本中的真实接触(1为接触,0为无接触)。我们将把不齐全样本数据中缺失的接触当做一个二项潜变量 $\eta_j$,其值为 $\eta_j = e_c, c = 1,2, e_1 = 1, e_2 = 0$。这两个样本的接触模型则为

$$\text{log it}[\Pr(X_{1j} = 1)] = \varrho_0$$

与

$$\text{log it}[\pi_1] = \varrho_0$$

其中,$\pi_1$ 为在不齐全数据样本中的对象属于第一个潜类的概率。

***测量模型***

在协变量测量误差问题上,经常需要设定一个假设,如在14.2.2节中,假设存在非差异性测量误差,也就是说给定真实接触的条件下,所测量的接触条件依赖于发病状态。设 $W_{1j}$ 与 $W_{0j}$ 分别表示校验样本与不齐全数据样本中测量到的接触。另外,设 $D_{1j}$ 与 $D_{0j}$ 分别表示校验样本与不齐全数据样本中的发病状态(1表示病例,0表示控制组)。对于校验样本的非差异性测量误差模型设定为

$$\text{log it}[\Pr(W_{1j} \mid X_{1j}, D_{1j})] = \alpha_0 + \alpha_1 X_{1j}$$

对于不齐全数据样本设定为

$$\text{log it}[\Pr(W_{0j} \mid \eta_j, D_{0j})] = \alpha_0 + \alpha_1 \eta_j$$

卡洛尔等(Carroll et al., 1993)指出,在校验样本中有证据表明存在差异性测量误差,因为校验样本中有部分案例所估计的灵敏度(18/23 = 0.78)高于控制组(16/32 = 0.50)。差异性测量误差饱和模型包括真实接触的效应、疾病状态及其交互项,即

$$\text{log it}[\Pr(W_{1j} \mid X_{1j}, D_{1j})] = \alpha_0 + \alpha_1 X_{1j} + \alpha_2 D_{1j} + \alpha_3 D_{1j} X_{1j}$$

与

$$\text{log it}[\Pr(W_{0j} \mid \eta_j, D_{0j})] = \alpha_0 + \alpha_1 \eta_j + \alpha_2 D_{0j} + \alpha_3 D_{0j} \eta_j$$

***疾病模型***

对于校验样本,在给定真实接触条件下,我们可设定一个对数概率回归模型,即

$$\text{log it}[\Pr(D_{1j} = 1 \mid X_{1j})] = \beta_0 + \beta_1 X_{1j}$$

对于剩余的样本,真实接触状态信息缺失,因而用潜类来表示。在不齐全数据样本

---

1 数据可从 gllamm.org/books 下载。

中，疾病状态 $D_{0j}$ 的模型则为

$$\text{logit}[\Pr(D_{0j}=1 \mid \eta_j)] = \beta_0 + \beta_1\eta_j$$

***结合模型***

用 $y_{ij}$ 表示该3个响应变量：接触（$i=1$）、测量（$i=2$）和疾病（$i=3$），以及虚拟变量 $d_{ri}=1$，当 $r=i$ 时，$d_{ri}=1$；否则，为0。用 $v_j=1$ 表示对象 $j$ 在校验样本中，否则 $v_j=0$。该结合模型（允许差异性测量误差）用GRC方程式可写为

$$\begin{aligned}\nu_{ij} &= d_{1i}[\varrho_0] + d_{2i}[\alpha_0 + \alpha_1 X_{1j}v_j + \alpha_1\eta_j(1-v_j) + \\ &\quad \alpha_2 D_{1j}v_j + \alpha_2 D_{0j}(1-v_j) + \alpha_3 X_{1j}D_{1j}v_j + \alpha_3\eta_j D_{0j}(1-v_j)] + \\ &\quad d_{3i}[\beta_0 + \beta_1 X_{1j}v_j + \beta_1\eta_j(1-v_j)] \\ &= \varrho_0 d_{1i} + \alpha_0 d_{2i} + \alpha_1 X_{1j}v_j d_{2i} + \alpha_2 D_{1j}v_j d_{2i} + \alpha_2 D_{0j}(1-v_j)d_{2i} + \\ &\quad \alpha_3 X_{1j}D_{1j}v_j d_{2i} + \beta_0 d_{3i} + \beta_1 X_{1j}v_j d_{3i} + \\ &\quad \eta_j[\alpha_1(1-v_j)d_{2i} + \alpha_3 D_{0j}(1-v_j)d_{2i} + \beta_1(1-v_j)d_{3i}]\end{aligned}$$

以及

$$\text{logit}[\pi_1] = \varrho_0$$

在此，$\alpha_1$，$\alpha_3$ 和 $\beta_1$ 为因子负荷，在校验样本中被约束为等于其对应的回归系数。

差异性测量误差模型与非差异性测量误差模型的参数估计结果如表14.3所示。根据测量模型的估计值，我们可估计出条件概率 $\Pr(W_j=1 \mid X_j, D_j)$，其结果如表14.4所示。

比较这两个模型，得到似然比值检验为5.14，2个自由度，因而相对简单的假设为非差异性测量误差的模型看上去更适合。假设为差异性测量误差模型的真实接触的对数优比估计为0.608（0.350），而假设为非差异性测量误差模型的对数优比估计值为0.958（0.237）。只使用校验样本所估计的优比为0.681（0.400）。请注意，使用全样本进行估计，我们仅在精度上得到非常细微的改进，这可能是由于低灵敏度及所测量的接触的特征的缘故。使用接触的不准确测量而非全样本的金标准，所得到的估计值为0.453（0.093）。正如我们所料，由于回归稀释的缘故，该估计与其他估计相比起来弱了很多。我们基于前瞻性似然的估计与卡洛尔等（Carroll et al.，1993）用回溯性似然的估计结果非常接近。对于不齐全数据，在给定所测量的接触及疾病状态条件下，我们可用真实接触的后验概率来估计对疱疹病毒的真实接触。这些后验概率，包含阳性预测值（PPV）以及阴性预测值（NPV）的信息，其结果如表14.5所示。

**表14.3　子宫颈癌数据估计值**

| | 差异性测量误差 | | 非差异性测量误差 | |
|---|---|---|---|---|
| 参数 | Est | (SE) | Est | (SE) |
| 接触模型 | | | | |
| $\varrho_0$[Cons] | -0.023 | (0.171) | 0.006 | (0.165) |
| 测量模型 | | | | |
| $\alpha_0$[Cons] | -0.791 | (0.259) | -1.063 | (0.224) |
| $\alpha_1$[TrueE] | 1.099 | (0.496) | 1.813 | (0.365) |
| $\alpha_2$[Case] | -0.665 | (0.616) | — | — |
| $\alpha_3$[TrueE]×[Case] | 1.649 | (0.955) | — | — |
| 疾病模型 | | | | |
| $\beta_0$[Cons] | -0.897 | (0.192) | -1.097 | (0.149) |
| $\beta_1$[TrueE] | 0.608 | (0.350) | 0.958 | (0.237) |
| 对数似然值 | -2 802.63 | | -2 805.19 | |

表 14.4 接触测量的条件概率 $\Pr(W_j=1|X_j,D_j)$

| $X_j$ [TrueE] | $D_j$ [Case] | 差异性测量误差 | 非差异性测量误差 |
|---|---|---|---|
| 1-特异性 | | | |
| 0 | 0 | 0.31 | 0.26 |
| 0 | 1 | 0.19 | 0.26 |
| 灵敏度 | | | |
| 1 | 0 | 0.58 | 0.68 |
| 1 | 1 | 0.79 | 0.68 |

表 14.5 真实接触的后验概率 $\Pr(X_j=1|W_j,D_j)$

| $W_j$ [MeasE] | $D_j$ [Case] | 差异性测量误差 | 非差异性测量误差 |
|---|---|---|---|
| 1-NPV | | | |
| 0 | 0 | 0.33 | 0.24 |
| 0 | 1 | 0.28 | 0.45 |
| PPV | | | |
| 1 | 0 | 0.59 | 0.65 |
| 1 | 1 | 0.86 | 0.83 |

## 14.4 职业培训与压力：服从者平均因果效应模型

### 14.4.1 介 绍

雷托和姚(Little and Yau, 1998)分析了维诺克尔等(Vinokur et al., 1995)所探讨的JOBS Ⅱ干预试验数据[1]。在过去13周内失业并且正在重新寻找工作的失业者被随机分成两组：一组给予5个半天的职业培训以及一个如何寻找工作的学习手册(试验组)，另一组只给予手册(控制组)。该干预的目标是为了预防精神健康问题以及促进高质量的再就业。在试验组中，存在不服从的问题，即其中有46%的人从不参加任何培训。

### 14.4.2 服从性问题

请考虑一下在一个随机对照研究中采用两种试验方法的情形。为了简单起见，我们将一组称为试验组，另一组称为控制组，在控制组中试验方法并不都是安慰剂。

英本斯和鲁宾(Imbens and Rubin, 1997ab)根据服从的状态将研究的参与者分成4种类型：**服从者**(compliers)(听从所分配的试验方法)，**一贯参与者**(always taker)(不管如何分配，总是采纳试验组的试验方法)，**从不参与者**(never-taker)(不管如何分配，从不采

1 该数据可从 gllamm.org/books 或密歇根大学社会科学研究中心的 ICPSR 网站 http://www.icpsr.umich.edu/access/index.html (在 JOBS#2739 下)下载。

纳试验组的试验方法）以及**抗逆者**（defiers）（总是采纳与所分配的相反的试验方法）。

重要的是，在平行试验中，服从的状态是无法被观测到的，因为一个对象是否采纳试验方法只能当他被分配到该组中才可能被观测到。首先考虑一下一个对象被分配到试验组的情形。如果该对象接受试验，那么他可能为服从者也可能是一贯参与者。如果他拒绝，那么他可能为从不参与者或抗逆者。接下来，对于一个对象被分配到控制组的情形，如果他接受试验，那么他可能为服从者或者为从不参与者。反之，则可能为一贯参与者或抗逆者。通过图 14.4 的图形表示可以直观地说明该现象。

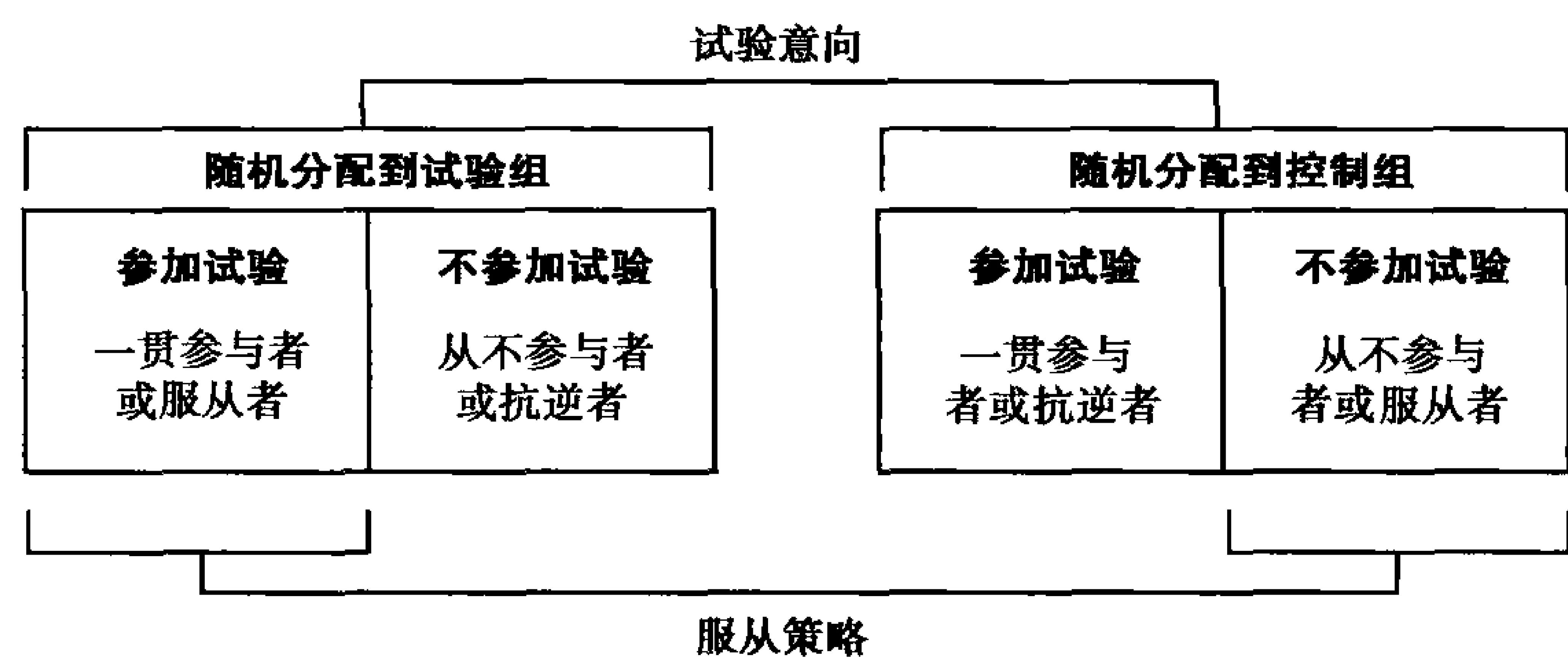



**图 14.4　服从状态以及不服从的几种方式**

通常来说，可以合理地假设不存在抗逆者，有时将此称为"单调"假设。这是因为通常在对照研究中，参与者都是自愿的；另外，不服从于试验分配通常是由于对于某个试验方法的先验偏好。因而，我们此后都假设不存在抗逆者。正如英本斯和鲁宾（Imbens and Rubin, 1997b）所证明的，该假设对于辨识非常有用。特别是，如果可以估计在两个组中实际接受所分配的试验的概率，就可以辨识出成为一贯参与者、从不参与者以及服从者的概率。

通过下面描述可以看到这一点（参看图 14.4 也将会很有帮助）：在试验组中，成为从不参与者的概率等于不接受试验的概率，因为不存在抗逆者。由于随机分配，成为从不参与者的概率与控制组中是相等的。在控制组中，成为服从者的概率则可通过接受**控制试验**（**对照试验**）的概率减去成为从不参与者的概率得到。由于随机分配，在试验组中成为服从者的概率与此相等。在试验组中，成为一贯参与者的概率可以通过接受试验的概率减去成为服从者的概率而得到。最后，随机分配保证了试验组中成为一贯参与者的概率与控制组的相等。

常规来讲，当存在不服从者时，有 3 种不同的分析方法：

（1）"**随机化分析**（as-randomized analysis）"或试验意向分析像图 14.4 那样比较分配到不同组的参与者的结果。试验意向效应即为试验分配的效应，而非接受试验分配的效应（通常称为"效果（effectiveness）"，与"功效（efficacy）"相对应）。通过随机试验分配，标准的试验意向估计量可避免选择性偏差（selection bias）的影响。

（2）"**实验分析**（as-treated analysis）"比较实际接受试验的参与者的结果。在此，随机分配原则被打破，对试验效应的因果解释可能是值得怀疑的。

（3）"**服从策略分析**（per-protocol analysis）"比较那些服从所分配的试验的那部分对象，但排除了那些背离所分配试验的对象，如图 14.4 所示。该分析无法用于说明个体因果效应。

英本斯和鲁宾（Imbens and Rubin, 1997b）介绍了另外一种试验效应，即"服从者平均因果效应（complier average causal effect, CACE）"。该效应是指真实服从者的试验效应，

即在试验组中的服从者与在控制组中的那部分如果被随机分配到试验组并成为服从者的那部分对象的平均结果差异。因而,CACE 可被看做是对功效的测量,而非对效果的测量。CACE 的形式化表示的关键在于区分"真实服从者"(在试验组和控制组中的服从者)与"观测到的服从者"(接受所分配的试验)。因而,CACE 建模的挑战在于对象的真实服从状态一般来说是未知的。形式上,服从者平均因果效应(标记为 $\delta_c$)被定义为

$$\delta_c \equiv \mu_{1c} - \mu_{0c}$$

其中,$\mu_{1c}$为试验组中的服从者结果均值,$\mu_{0c}$为控制组中的服从者结果均值。

在这个 JOBS Ⅱ 干预中,对于那些被分配到控制组中的人来说,职业培训干预是不可能的。因而,其中就存在一贯参与者,即如果不存在分配,他们都将选择参与职业培训。然而,我们无法估计成为一贯参与者的概率,因为在控制组中的一贯参与者无法表现出来,并且在试验组中的一贯参与者与服从者也无法区分。这个问题的一个解决方法是将一贯参与者和服从者合并为一个单独的组(如 Little and Yau, 1998),或者假设不存在一贯参与者(如 Jo, 2002)。我们采用前者,并将这个服从者与一贯参与者的合并组宽泛地称为"服从者"。最后所得到的"服从者"平均因果效应仍然是有意义的,因为它表示如果给予机会并会参加培训的那些对象的培训效应。JOBS Ⅱ 干预的设计以及在此情形下,CACE 的含义如图 14.5 所示,其中短画线说明,在控制组中我们不清楚如果给予机会的话,谁将参加试验,谁不参加。

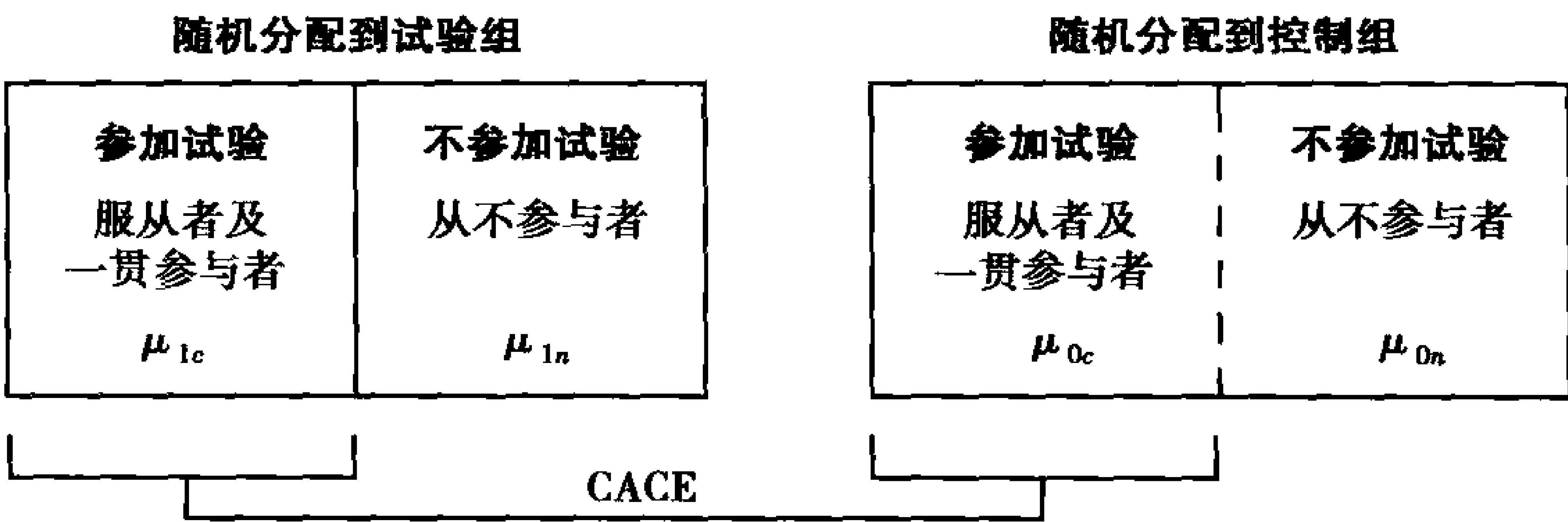



图 14.5 职业培训对照研究的服从者平均因果效应(CACE)

我们还假设了在试验组和控制组中,从不参与者的均值相等,即

$$\mu_{0n} = \mu_{1n}$$

(通常称为"排斥约束(exclusion restriction)")。

### 14.4.3 CACE 建模

阿奎斯提等(Angrist et al., 1996)、英本斯和鲁宾(Imbens and Rubin, 1997ab)、雷托和姚(Little and Yau, 1998)以及乔(Jo, 2002)都讨论过 CACE 模型。穆森(Muthén, 2002)用此职业培训数据将该模型形式化为潜类模型,我们将采用该方法。

***服从模型***

设 $c_j$ 为服从者(或一贯参与者)对比从不参与者的虚拟变量,设 $r_j$ 为被随机分配到试验组或控制组的虚拟变量。

在控制组中,服从的状态不可观测,因而用 $\eta_j$ 表示,$\eta_j$ 为离散潜变量,其值分别为 $e_1=1, e_2=0$。由于随机分配,在此模型中,我们可以设置控制组中潜在服从的参数等于试验组中观测到的服从的参数。

$$\log\text{it}[\pi_{1j}] \equiv \log\text{it}[\Pr(c_j=1 \mid r_j=0)] = \mathbf{w}_j'\boldsymbol{\varrho} = \log\text{it}[\Pr(c_j=1 \mid r_j=1)]$$

服从模型的协变量如下:

- [Age] 按年份表示的年龄。
- [Motivate] 参与的动机。
- [Educ] 教育程度。
- [Assert] 自信程度。
- [Single] 单身,虚拟变量。
- [Econ] 经济压力。
- [Nonwhite] 非白人,虚拟变量。

***压力模型***

所关心的响应变量为从研究开始到培训 6 个月后压力得分(霍普金斯症量表的 11 题器子量表)的变化。如果已知控制组中的服从,则可将压力模型转化为

$$y_j = \beta_0 + \beta_1 c_j(1 - r_j) + \beta_2 c_j r_j + \epsilon_j$$

其中,$\epsilon_j \sim N(0,\theta)$使得$\mu_{0n} = \mu_{1n} = \beta_0, \mu_{0c} = \beta_0 + \beta_1, \mu_{1c} = \beta_0 + \beta_2$ 以及 $\delta_c = \beta_2 - \beta_1$。然而,在第二项(控制组)中的 $c_j$ 不可能被观测到。因此,我们用潜在服从 $\eta_j$ 的形式将模型改写为

$$y_j = \beta_0 + \beta_1 \eta_j(1 - r_j) + \beta_2 c_j r_j + \epsilon_j$$

通过设定

$$y_j = \beta_0 + \mathbf{x}_j'\boldsymbol{\alpha} + \beta_1 \eta_j(1 - r_j) + \beta_2 c_j r_j + \epsilon_j$$

我们可以将在控制组中具有常数效应 $\boldsymbol{\alpha}$ 的协变量 $\mathbf{x}_j$ 包含进来。

对于该压力模型,所考虑的协变量如下:

- [Basedep] 基准压力得分。
- [Risk] 基准风险得分;是关于压力、金融压力以及自信程度的综合量表。

***联合模型***

在 CACE 建模中,我们有两个响应变量,试验组中的服从 $c_j$ 以及全部样本中的压力 $y_j$。我们将这两个响应变量分别用 $y_{1j}$与 $y_{2j}$来表示,并且定义其对应的虚拟变量,当 $r = i$ 时,$d_{ri} = 1$;反之为 0。响应模型用 GRC 方程式可重写为

$$\nu_{ij} = d_{i1}\mathbf{w}_j'\boldsymbol{\varrho} + \beta_0 d_{2i} + d_{2i}\mathbf{x}_j'\boldsymbol{\alpha} + \eta_j\beta_1(1 - r_j)d_{2i} + \beta_2 c_j r_j d_{2i}$$

其中,当 $i = 1$ 时,设定一个对数概率链接函数和伯努利分布;当 $i = 2$ 时,设定一个同一性链接函数和正态密度分布。该结构模型就为

$$\log\text{it}[\pi_{1j}] = \mathbf{w}_j'\boldsymbol{\varrho}$$

在此,我们将重复雷托和姚(Little and Yau, 1998)关于“高风险”组的分析。该组包含 335 个被随机分配到职业培训的对象,另有 167 个被随机分配到控制组的对象。在 335 个分配到职业培训的对象中,仅有 183 个(55%)实际上参加了培训。该 CACE 模型的参数估计以及标准误结果如表 14.6 所示。

年龄高、动机强以及高教育程度与服从高度关联,而高自信程度则会降低概率。在这些参加了职业培训的人中,似乎存在压力削减的现象。当没有包含协变量时,服从的平均因果效应估计值为 -0.14 (0.14)。当包含协变量时,服从的平均因果效应更加显著,为 -0.31 (0.12)。做个对比,与 CACE 模型一样,包含了协变量基准压力以及基准风险的用常规方法估计的试验效应如下:

(1)试验意向效应:-0.15 (0.07)。

(2)试验效应(包括控制组中的非参与者):-0.18 (0.07)。

(3)服从策略效应(排除非参与者):-0.21 (0.08)。

正如所期望的,试验意向效应最低,因为将非参加者包含到试验组中冲减了试验效

应。对比 CACE 模型以及服从策略效应也非常有意思，因为它们均将从不参加者排除出试验组，但 CACE 模型还将这部分人从控制组中排除掉。如果从不参加者存在微弱的试验效应，服从策略效应将会更高，但令人奇怪的是，在本例子中并没有发生($\hat{\beta}_1$ 为正数)。

**表 14.6　CACE 模型参数估计值，服从模型包含或未包含协变量**

| 参数 | 无协变量 | | 协变量 | |
|---|---|---|---|---|
| | Est | SE | Est | SE |
| 服从模型 | | | | |
| $\varrho_0$[Cons] | -0.19 | (0.11) | -8.74 | (1.58) |
| $\varrho_1$[Age] | | | 0.08 | (0.01) |
| $\varrho_2$[Motivate] | | | 0.67 | (0.16) |
| $\varrho_3$[Educ] | | | 0.30 | (0.07) |
| $\varrho_4$[Assert] | | | -0.38 | (0.15) |
| $\varrho_5$[Single] | | | 0.54 | (0.28) |
| $\varrho_6$[Econ] | | | -0.16 | (0.16) |
| $\varrho_6$[Nonwhite] | | | -0.50 | (0.31) |
| 压力模型 | | | | |
| $\beta_0$[Cons] | -0.39 | (0.07) | 1.63 | (0.28) |
| $\beta_1$ | 0.02 | (0.17) | 0.18 | (0.13) |
| $\beta_2$ | -0.12 | (0.09) | -0.13 | (0.08) |
| $\delta_c=\beta_2-\beta_1$ | -0.14 | (0.14) | -0.31 | (0.12) |
| $\alpha_1$[Basedep] | | | -1.46 | (0.18) |
| $\alpha_1$[Risk] | | | 0.91 | (0.26) |
| $\theta$ | 0.60 | (0.04) | 0.51 | (0.03) |
| 对数似然值 | -815.15 | | -729.41 | |

## 14.5　医师建议及饮酒：内生试验模型

### 14.5.1　介　绍

酒精滥用已经是一个公众关注的健康问题，它不仅带来健康问题，而且还带来诸如与饮酒相关的交通意外等。为了降低饮酒相关的问题，其中一种方式是通过医师劝诫问题饮酒者少饮酒。医师建议对降低饮酒的功效已经在有临床对照试验中证明过。然而，基于观测数据的研究还是必须的，因为在日常活动中，功效(efficacy)并不一定转化为效果(effectiveness)。

肯科和特尔扎(Kenkel and Terza, 2001) 分析了 1990 年国家健康访谈抽样调查的核

心问卷以及附录问卷的数据。这个数据[1]由 2 467 个男性子样本构成,这些男性都是饮酒者并被告知患有高血压。响应变量为在过去两周中所饮用的酒精饮料数量,即事件数变量。28%的饮酒者报告他们曾被建议过减少喝酒。该研究的目的是估计医师建议的"试验效应"。

与随机试验研究相反,根据观测数据来估计试验效应的一个主要问题在于试验通常是"内生的",也就是说,试验与未观测到的异质性相关。例如,一个预后不良的对象(未观测到的异质性)很可能自我选择到试验组中并努力做到最好。在这种情况下,毫无疑问,常规的模型估计将得到严重偏误的试验效应估计。因此,在计量经济学中(如 Heckman, 1978),提出了试图纠正选择性偏误的模型,这些模型通常被称为内生试验模型(endogenous treatment models)。在本节,我们应用特尔扎及其同事所讨论的内生试验模型来分析事件数响应变量(如 Terza, 1998; Kenkel and Terza, 2001)。

## 14.5.2 饮酒模型

主要的目标是分析如下变量的效应:

- [Advice]对问题"是否有医师建议您减少饮酒?"的回答,处理成虚拟变量 $T_j$。

另外,对饮酒过程 $\mathbf{x}_j$ 所包含的协变量如下:

- [HiEduc] 虚拟变量,是否受过高等教育(超过 12 年对应于小于或等于≤12 年)。
- [Black] 虚拟变量,种族(黑人对应于非黑人)。

为了简单起见,这里所展现的只是肯科和特尔扎(Kenkel and Terza, 2001)的分析所使用的协变量中的一小部分。

由于消费量 $y_j$ 为事件数变量,很自然,我们会考虑泊松回归模型

$$\Pr(y_j;\mu_j) = \frac{\mu_j^{y_j}\exp(-\mu_j)}{y_j!}$$

其中,期望值 $\mu_j$ 被结构化为对数线性模型

$$\log(\mu_j) = \alpha T_j + \mathbf{x}_j'\boldsymbol{\beta}$$

请注意,在这个模型中,[Advice]的试验效应用 $\alpha$ 表示。该模型的估计值如表 14.7 所示的第二栏。由于零响应出现的比例占到了 21%,我们考虑使用能诱发过度离散及使零值数目膨胀的模型。第 11.2 节所讨论的零堆积泊松模型看起来并不适用于这个例子,因为戒酒者被排斥在样本之外。因而,我们考虑使用随机截距泊松回归来对过度离散建模,即

$$\log(\mu_j) = \alpha T_j + \mathbf{x}_j'\boldsymbol{\beta} + \zeta_j$$

其中,$\zeta_j \sim \mathrm{N}(0,\psi)$。该模型的估计值如表 14.7 所示的第三栏。

对于这两个泊松模型,[HiEduc] 与 [Black] 对于饮酒都存在负效应,这看上去不合情理。然而,关键的值得指出的一点在于,[Advice]对于饮酒具有相当不恰当的正效应!除此之外,当包含其他潜在的混淆因素时,符号的方向也并不改变(见 Kenkel and Terza, 2001)。如果我们根据这些结果,那么就应该放弃对问题饮酒者的医师建议。请注意,存在一个非常明显的过度离散,其估计的随机截距方差为 2.90。

---

1 该数据可从《应用计量经济学期刊》的网站数据库下载,地址为 http://qed.econ.queensu.ca/jae/2001-v16.2/kenkel-terza/。

表 14.7 饮酒模型及建议模型的估计值及标准误

| | 泊松模型 | | 过度离散泊松模型 | | 正态概率模型 | | 内生实验模型 | |
|---|---|---|---|---|---|---|---|---|
| 试验 | Est | (SE) | Est | (SE) | Est | (SE) | Est | (SE) |
| 固定部分 | | | | | | | | |
| 饮酒模型 | | | | | | | | |
| $\alpha$[Advice] | **0.47** | (0.01) | **0.59** | (0.08) | | | **-2.42** | (0.23) |
| $\beta_0$[Cons] | 2.65 | (0.01) | 1.43 | (0.06) | | | 2.32 | (0.09) |
| $\beta_1$[HiEduc] | -0.18 | (0.01) | 0.02 | (0.07) | | | -0.29 | (0.10) |
| $\beta_2$[Black] | -0.31 | (0.02) | -0.29 | (0.11) | | | 0.20 | (0.11) |
| 建议模型 | | | | | | | | |
| $\gamma_0$[Cons] | | | | | -0.48 | (0.08) | -1.13 | (0.16) |
| $\gamma_1$[HiEduc] | | | | | -0.25 | (0.06) | -0.40 | (0.10) |
| $\gamma_2$[Black] | | | | | 0.30 | (0.08) | 0.60 | (0.15) |
| $\gamma_3$[HlthIns] | | | | | -0.27 | (0.07) | -0.33 | (0.10) |
| $\gamma_4$[RegMed] | | | | | 0.18 | (0.07) | 0.39 | (0.10) |
| $\gamma_5$[Heart] | | | | | 0.17 | (0.08) | 0.51 | (0.11) |
| 随机部分 | | | | | | | | |
| 方差 | | | | | | | | |
| $\psi$ | | | 2.90 | (0.11) | | | 2.50 | (0.69) |
| 负荷 | | | | | | | | |
| $\lambda$ | | | | | | | 1.43 | (0.15) |
| 对数似然值 | -32 939.15 | | -8 857.85 | | -1 419.90 | | -10 254.02 | |

### 14.5.3 建议模型

为了获得关于建议过程的更多洞见，对于试验[Advice]，我们设定并估计了一个正态概率模型。将该模型形式化为潜响应模型将非常有帮助，即

$$T_j^* = \mathbf{w}_j'\boldsymbol{\gamma} + \epsilon_j$$

其中，$\epsilon_j \sim \mathrm{N}(0,1)$。试验 $T_j$ 根据阈值模型而得

$$T_j = \begin{cases} 1 & 当\ T_j^* > 0 \\ 0 & 其他 \end{cases}$$

协变量向量 $\mathbf{w}_j$ 除了包括前面已经介绍的协变量外，还包含 3 个利用健康服务的二项变量：

- [HlthIns] 虚拟变量，是否有健康保险。
- [RegMed] 虚拟变量，是否有注册的医疗保健。
- [Heart] 虚拟变量，心脏健康状况。

另外，为了简单起见，我们还是只包括了肯科和特尔扎(Kenkel and Terza, 2001)所使用的协变量的一部分。

该模型的估计结果如表 14.7 的第三栏，其中，我们可以看到[Black]、[RegMed]及[Heart]对[Advice]具有正效应，而[HiEduc]与[HlthIns]则具有负效应。

## 14.5.4　饮酒及建议联合模型：建议作为内生试验

对于饮酒以及建议过程，很可能存在共享的未观测到的异质性，因为医师可能更倾向于向那些存在风险的患者提出建议。因此，我们分析下面这个关于饮酒及建议的联立模型，即为内生试验模型

$$\begin{aligned}\log(\mu_j) &= \alpha T_j + \mathbf{x}_j'\boldsymbol{\beta} + \lambda\zeta_j \\ T_j^* &= \mathbf{w}_j'\gamma + \zeta_j + \epsilon_j\end{aligned}$$

其中，$\zeta_j \sim \mathrm{N}(0,\psi)$表示共享的未观测到的异质性因子，$\lambda$ 为因子负荷，$\epsilon_j \sim N(0,1)$。请注意，该因子模型只是对特尔扎（Terza，1998）模型的参数优化，将积分维度从 2 维削减为 1 维。

根据该内生试验模型，可得

$$\mathrm{Var}[\log(\mu_j) \mid T_j,\mathbf{x}_j] = \lambda^2\psi \geqslant 0$$

从而，允许过度离散

$$\mathrm{Var}[T_j^* \mid \mathbf{w}_j] = \psi + 1$$

与

$$\mathrm{Cov}[\log(\mu_j), T_j^* \mid \mathbf{w}_j] = \lambda\psi$$

请注意，由于正态概率模型中增加的残方差，系数 $\boldsymbol{\gamma}$ 应该增加约$\sqrt{\psi+1}$。将饮酒的响应 $i=1$ 与建议的响应 $i=2$ 看做为在患者 $j$ 内聚合，我们对于饮酒定义虚拟变量 $d_{1i}=1$，对于建议定义虚拟变量 $d_{2i}=1$。使用 GRC 方程式，内生试验模型可重写为

$$\begin{aligned}\nu_{ij} &= d_{1i}[\alpha T_j + \mathbf{x}_j'\beta + \lambda\zeta_j] + d_{2i}[\mathbf{w}_j'\gamma + \zeta_j] \\ &= \alpha d_{1i}T_j + d_{1i}\mathbf{x}_j'\beta + d_{2i}\mathbf{w}_j'\gamma + \zeta_j[d_{1i}\lambda + d_{2i}]\end{aligned}$$

对于 $i=1$，为对数链接函数以及泊松分布；对于 $i=2$，为正态概率链接函数。该模型的路径图如图 14.6 所示，其中，$\mathbf{w}_j'=(\mathbf{x}_j',\mathbf{z}_j')$。

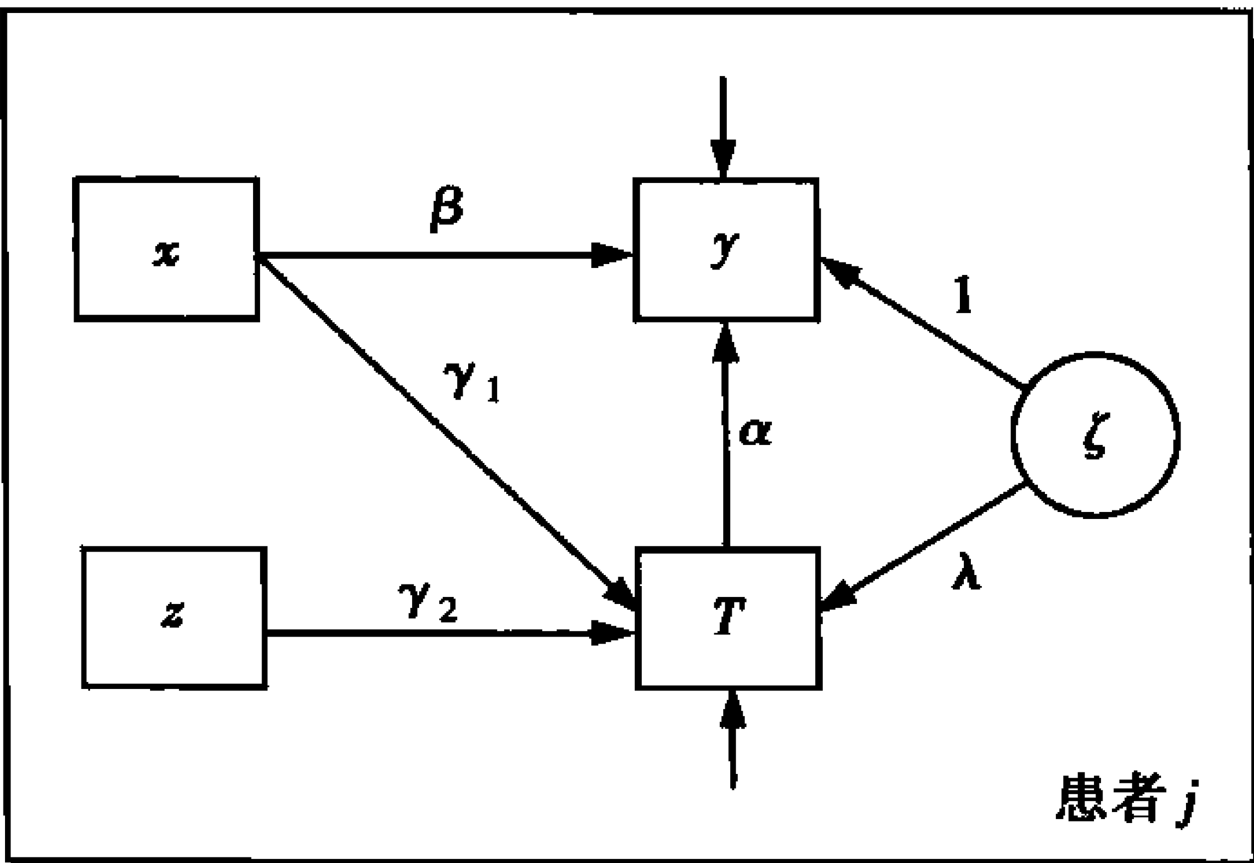



**图 14.6　内生试验模型路径图**

建议模型的一个很重要的特点在于 $T_j^*$ 与 $\zeta_j$ 相互独立，这也意味着在饮酒模型中，$T_j$ 与 $\zeta_j$ 相互独立。因此，在饮酒模型中，[Advice]就为“内生”变量，并且，对试验效应 $\alpha$ 的有效推论也必须基于这个同时包括饮酒以及建议作为响应变量或“内生”变量的联立模型。相反，约束 $\lambda=0$ 将似然分解为对于饮酒的常规泊松模型及对于建议的正态概率模型两项（如果 $\psi=0$，与上述分析等价）。

重要的是，将患者随机分配给[Advice]也就意味着试验是外生的，因为 $T_j$ 就将独立于未观测到的异质性 $\zeta_j$。在这种情况下，[Advice]成为“外生的”，并且对于 $\alpha$ 的有效推

论可以从常规泊松回归模型得到。可以看到，$\lambda$ 决定了协方差的符号，并且当 $\lambda=1$ 时，可得到共享的随机截距模型（对于饮酒及建议）。

**生活规律转移模型**（switching regime model）即为内生试验的一般化模型。在此，用一个正态概率模型决定患者被分配到试验组或非试验组，即

$$y_j^* = \mathbf{w}_j'\gamma + \zeta_{1j} + \zeta_{2j} + \epsilon_j$$

其中，$\zeta_{1j} \sim N(0,\psi_{11})$ 且 $\zeta_{2j} \sim N(0,\psi_{22})$，独立分布且独立于 $\epsilon_j$。在试验方案中，饮酒可用模型表示为

$$\log(\mu_{2j}) = \beta_{20} + \mathbf{x}_j'\boldsymbol{\beta}_2 + \lambda_2\zeta_{2j}$$

在非试验方案中，则为

$$\log(\mu_{1j}) = \beta_{10} + \mathbf{x}_j'\boldsymbol{\beta}_1 + \lambda_1\zeta_{1j}$$

其中，$\mathbf{x}_j$ 不再包含常数项。试验效应现在就成为 $\alpha=\beta_{20}-\beta_{10}$。注意，患者只可能被分配到其中一种方案，而不可能两者兼有，从而方案中的响应则表示潜响应。不幸的是，在本例中，对于这个方案转移模型无法得到可靠的估计。

该内生试验模型的估计值如表 14.7 所示的第四栏。需要重点注意的是医师建议的效应现在逆转了。获得建议将导致降低饮酒，与临床对照试验的结果一致。引入更多的协变量（见 Kenkel and Terza, 2001）不会改变这个结论。由于 $\hat{\lambda}$ 为正数，看起来在饮酒与建议过程的未观测到的异质性中存在正相关关系。因此，在控制了协变量（观测到的异质性）之后，重度饮酒的患者更可能得到关于饮酒的建议。正如所期望的，包括未观测到的异质性使建议模型的估计值进一步增大。

请注意，由于在饮酒模型中，关于健康服务利用的变量的效应被隐含地设置为零，从而内生试验模型存在一个所谓的排斥约束（与 CACE 的排斥约束不同）。根据肯科和特尔扎（Kenkel and Terza, 2001），这些约束可从将酒精需求当做消费品的标准模型引申而来。尽管有助于模型辨识，但对模型辨识来说，这些约束并不是必需的。释放这些约束并不会改变主要的结果。反之，则在解释结果时需要极其小心。

请注意，可得到一个事件数的样本选择模型作为此内生试验模型的一个特殊例子，其中 $\alpha=0$，并且 $T_j$ 为样本选择的标识变量，当对象 $j$ 在样本中时其值为 1，反之为 0。该模型也就是赫克曼针对连续响应变量所提出的样本选择模型的一般化（如 Heckman, 1979）。

## 14.6 肝硬化治疗：生存与指标联合模型

### 14.6.1 介 绍

安德森等（Andersen et al., 1993）分析了丹麦的一个确诊为肝硬化（通常由于酒精滥用引起）患者的随机对照试验数据。这些患者被随机分配到荷尔蒙强的松试验组或安慰剂控制组。根据更严格的标准重新评估了 488 位患者初始的活组织切片检查，其中 251 位分配给强的松，另 237 位分配给安慰剂[1,2]。强的松试验组标记为[Treat]，其对应的虚拟变量就用 $T$ 表示。

---

1 我们感谢安德森（Per Kragh Andersen）提供该数据。

2 数据可从 gllamm.org/books 下载。

该对照试验的主要目的是确定强的松是否能够降低这些硬化患者的死亡风险。如果患者没能继续获得跟踪或者在观测期结束时仍然生存,那么就被认为是删失的。除此之外,数据还包括凝血酶原、肝功能生化指标的重复测量。凝血酶原指标是基于肝脏所产生的凝血因子Ⅱ,Ⅶ及Ⅹ的血液测试而得(我们将原始指标除以10,以避免非常出现微小的回归系数)。我们还考虑到了指标的二进制处理,即如果原指标值高于70则定义为正常,反之为不正常。

对凝血酶原的测量原计划定在随机分配之后的第3、第6和第12个月进行,再往后每年进行一次,但实际追踪的次数是不规则的。因此,数据的一个问题在于高度不均衡的设计以及指标的测量缺失。

关于存活,还存在另外一个问题,即对于每个存活超过本次测量的对象,当其测量缺失时,必须获得它的时变协变量(见2.4.1节风险膨胀数据的图2.1)。然而,在每次测量缺失时,对指标的测量也是不可能的,因此,就必须使用某种插值方法。克里斯坦森等(Christensen et al.,1986)做了一个不切实际的假设,他们认为在两次追踪之间,时变协变量是常量。与他们相反,我们建议使用增长曲线模型。

另外,看起来似乎肝功能也可能是一个干预变量,这意味着强的松对于死亡风险具有通过肝功能的间接效应,以及直接的试验效应。分解这些效应首先可能有助于我们理解试验如何产生效果。其次,它可能能够揭示指标能否被当做生存的替代。如果不存在直接效应,控制了指标后,生存条件依赖于试验。因此,一旦指标已知,关于生存的信息并不包含试验效应信息以外的其他任何信息,指标被称为"完美替代"(如Prentice,1989)。那么,指标就可替代生存当做响应变量,这也带来很多操作上的便利,如可以减少所需要的跟踪次数。

为了分析这个问题,我们设定一个结构方程生存模型,包括一个潜在指标模型及一个包含一个潜在协变量的风险模型。

### 14.6.2 潜在指标模型

设$t_{ij}$为患者$j$在第$i$次测量的时点。在时点$t_{ij}$所观测到的指标用$y_{ij}$表示,并通过测量模型与潜变量或"真实的"指标$\eta_{ij}$相关。

$$y_{ij} = \beta_0 + \eta_{ij}^{(2)} + \epsilon_{ij}$$

其中,$\epsilon_{ij} \sim \mathrm{N}(0,\theta)$。

潜在指标的结构模型可设定为

$$\eta_{ij}^{(2)} = \gamma_1 t_{ij} + \gamma_2 T_j + \eta_j^{(3)} \tag{14.2}$$

其中,$\eta_j^{(3)} \sim \mathrm{N}(0,\psi)$独立分布于$\epsilon_{ij}$。请注意,在结构模型中不存在误差项$\zeta_{ij}^{(2)}$,或者说其对应的$\mathrm{Var}(\zeta_{ij}^{(2)})=0$。另外,我们还考虑了一个更一般化的结构模型,除了随机截距外,包含一个关于时间的随机斜率(二元正态随机截距和斜率),但是对数似然值并没有任何增加。

将其代入测量模型,替换潜在指标$\eta_{ij}^{(2)}$,化约后的指标模型就成为

$$y_{ij} = \beta_0 + \gamma_1 t_{ij} + \gamma_2 T_j + \eta_j^{(3)} + \epsilon_{ij}$$

该模型为随机截距线性增长模型。请注意,$\epsilon_{ij}$通常被解释为测量误差(如Faucett and Thomas, 1996; Wulfsohn and Tsiatis, 1997; Xu and Zeger, 2001),但这是有问题的,因为该项混合了真实指标与线性时间趋势的偏差。这两个误差项之所以不能被拆解,是因为在每个时间点,对每个患者仅有一次测量。

对于二项指标变量“正常凝血酶原与不正常凝血酶原”，我们根据凝血酶原指标的70%切割点建立一个正态概率模型。该指标模型与上面介绍的模型相似，潜响应 $y_{ij}^*$ 替代了 $y_{ij}$。二项测量根据下面的阈值模型生成

$$y_{ij} = \begin{cases} 1 & 当\ y_{ij}^* > 0 \\ 0 & 其他 \end{cases}$$

### 14.6.3 包含潜在协变量的风险模型

设 $t_{rj}$ 为患者 $j$ 所存活的第 $r$ 次死亡时点，在时点 $t_{rj}$ 的风险用 $h_{rj}$ 表示。包含潜在指标 $\eta_{rj}^{(2)}$（在时点 $t_{rj}$）为协变量的比例风险模型可设定为

$$\ln h_{rj} = \ln h_{rj}^0 + \lambda \eta_{rj}^{(2)} + \alpha_4 T_j$$

其基准的风险模型为

$$\ln h_{rj}^0 = \alpha_0 + \alpha_1 t_{rj} + \alpha_2 t_{rj}^2 + \alpha_3 t_{rj}^3$$

根据系统地检查三阶分段多项式（third degree fractional polynomials）（如 Royston and Altman，1994），选择该三次多项式模型。然后，我们将潜在指标插值到必需的死亡时点 $t_{rj}$ 中，并替换结构指标模型（14.2）中的 $\eta_j^{(2)}$，得到化约的风险模型

$$\begin{aligned} \ln h_{rj} &= \ln h_{rj}^0 + \lambda(\gamma_1 t_{rj} + \gamma_2 T_j + \eta_j^{(3)}) + \alpha_4 T_j \\ &= \ln \bar{h}_{rj}^0 + [\lambda\gamma_2 + \alpha_4] T_j + \lambda \eta_j^{(3)} \end{aligned}$$

其中

$$\ln \bar{h}_{rj}^0 = \alpha_0 + [\lambda\gamma_1 + \alpha_1] t_{rj} + \alpha_2 t_{rj}^2 + \alpha_3 t_{rj}^3$$

为化约的基准风险。

我们可以看到，对数风险的总体试验效应 $\lambda\gamma_2 + \alpha_4$ 被分解为通过潜在指标的间接效应 $\lambda\gamma_2$ 以及直接效应 $\alpha_4$。如果不存在直接效应（$\alpha_4 = 0$），那么给定指标，对数风险条件依赖于试验。在这种情况下，如果指标已知，那么生存信息并不包含关于试验效应的其他附加信息，且指标 $\eta_{ij}^{(2)}$ 为“完美替代”。

该模型的路径图如图 14.7 所示，其中，字符 $i$，$r$ 和 $j$ 表示在每个框中变动的变量下标，$D_{rj}$ 为虚拟变量，当患者 $j$ 在风险集 $r$ 死亡为 1，反之为 0（生存模型的响应变量）。

连续指标模型和二项指标模型的参数估计和标准误如表 14.8 所示。注意，由于 $\theta$ 在正态概率指标模型中无法辨识，在估计连续指标模型时，我们对 $\theta$ 进行了约束，以使模型参数估计和标准误的比较更为简单。对于指标模型，我们注意到连续指标模型和二项指标模型的参数估计非常近似。但是[Cons]变量例外，它在两个模型中的角色不同，其在二项模型中与阈值紧密相关。正如所预期的，由于在二项处理中丢失了部分信息，从而正态概率模型估计的标准误更高。这些估计结果表明指标随着时间增高，但是[Treat]降低指标。连续观测的指标和二项指标的潜变量的组间残相关分别为 0.52 和 0.58。

至于风险模型，无论是否将观测的指标二进制化，参数估计和标准误都非常接近。除了[Treat]变量，其估计值的绝对值在二项模型中更小。对肝功能（潜在指标）所估计的试验效应为 $\hat{\gamma}_2 = -0.64$，其 95% 置信区间为（$-1.00$，$-0.28$）。因此，荷尔蒙强的松对肝功能有负面的“副作用”。所估计的肝功能对死亡风险的效应为 $\hat{\lambda} = -0.38$，其 95% 置信区间为（$-0.46$，$-0.31$）。对应的风险比（hazard ratio）为 $\exp(\hat{\lambda}) = 0.68$。正如所预期的，正常的肝功能降低死亡风险。所估计的对死亡风险的直接试验效应为 $\hat{\alpha}_4 = -0.18$（风险比 0.84），95% 置信区间为（$-0.42$，$0.06$）。因此，试验的直接效应看上去降低了

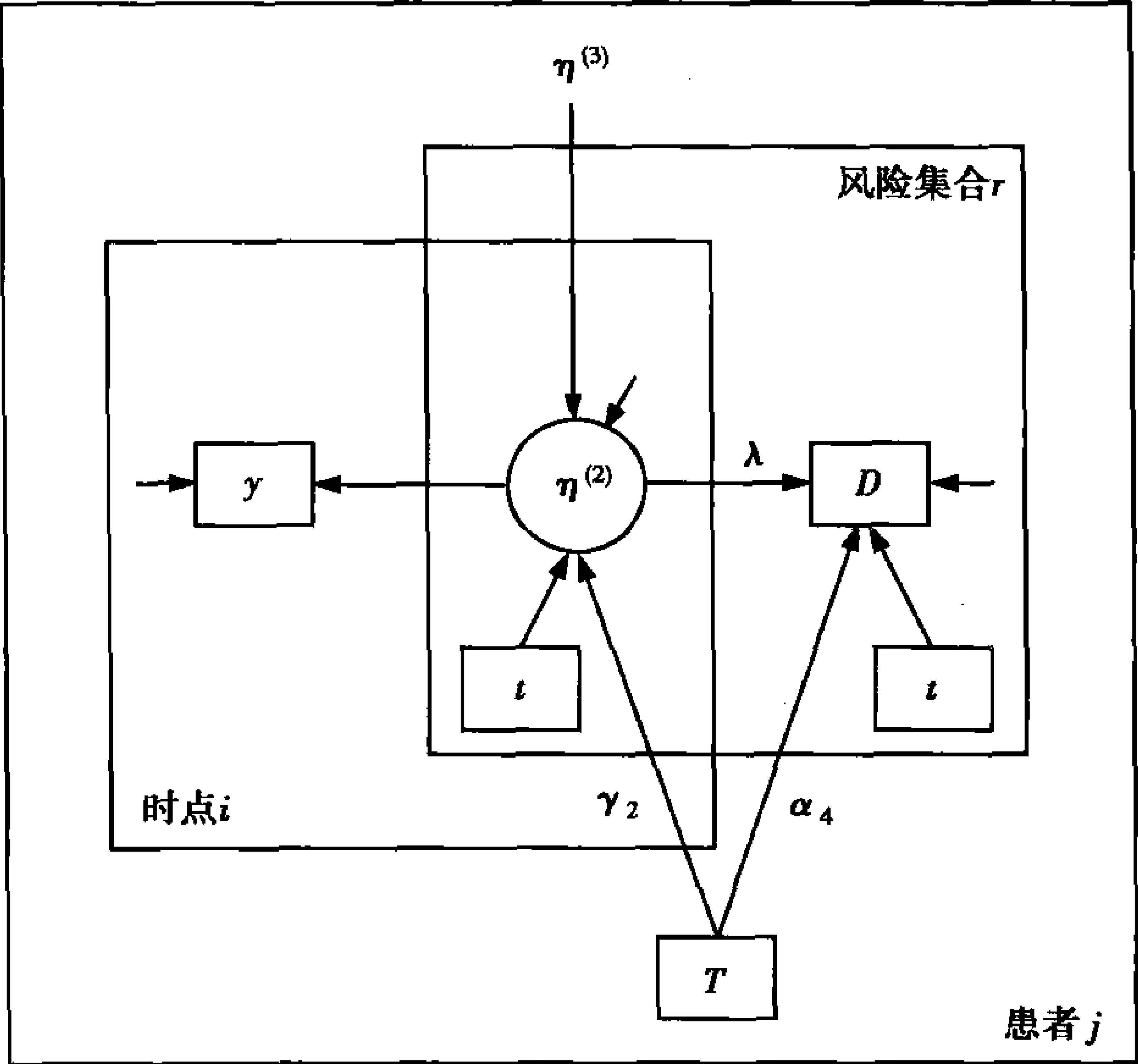



图 14.7 结构方程风险模型路径图

一些风险，但该估计是不够精确的。试验的间接效应估计值为 $\hat{\lambda}\hat{\gamma}_2=0.25$（风险比 1.28），95%置信区间为(0.08,0.41)。死亡风险的增加是由于试验对肝功能的副作用并进而增加风险而引起。总体的试验效应，即直接效应和间接效应的加总，其估计值为 $\hat{\lambda}\hat{\gamma}_2+\hat{\alpha}_4=0.07$（风险比 1.07），95%置信区间为(−0.22,0.35)。因此，没有证据证明强的松对肝硬化患者的死亡风险具有帮助作用。有趣的是，强的松从此以后不再作为治疗肝硬化的手段。

至于数据替代，直接效应 $\alpha_4$ 在5%水平上并没有显著区别于0，指标很可能被认为是一种较为完美的替代。然而，由于指标的间接效应是有害的，因此这是没有意义的。甚至，如果我们的目的是估计对生存的试验效应的大小，而非仅仅估计是否有试验效应的话，即使是最为完美的替代也不如生存期的信息量大。

表 14.8 结构方程风险模型估计值（连续指标模型和二项指标模型）

| | 连续模型 | | 二项模型 | |
|---|---|---|---|---|
| | Est | (SE) | Est | (SE) |
| 指标模型 | | | | |
| $\beta_0$[Cons] | 7.88 | (0.12) | 0.84 | (0.19) |
| $\gamma_1$[Time] | 0.15 | (0.02) | 0.11 | (0.03) |
| $\gamma_2$[Treat] | −0.64 | (0.19) | −0.57 | (0.27) |
| $\psi$ | 3.71 | (0.26) | 4.66 | (0.50) |
| $\theta$ | 3.43 | (0.97) | 3.43 | (—) |
| 风险模型 | | | | |
| $\alpha_0$[Cons] | −7.96 | (0.12) | −7.97 | (0.12) |
| $\alpha_1$[Time] | −0.15 | (0.05) | −0.16 | (0.05) |
| $\alpha_2$[Time]$^2$ | 0.10 | (0.02) | 0.10 | (0.02) |
| $\alpha_3$[Time]$^3$ | −0.01 | (0.00) | −0.01 | (0.00) |
| $\alpha_4$[Treat] | −0.18 | (0.12) | −0.10 | (0.13) |
| $\lambda$ | −0.38 | (0.04) | −0.34 | (0.04) |
| 对数似然值 | −8 422.87 | | −3 527.28 | |

## 14.7 总结与进一步阅读

我们在前瞻性和回溯性研究中讨论了连续和离散真实协变量的测量误差模型，并采用了重复测量和校验样本的例子。卡洛尔等（Carroll et al., 1995a）以及古斯塔夫松（Gustafson, 2004）讨论了协变量测量误差模型。罗德尔等（Roeder et al., 1996）、沙弗尔（Schafer, 2001）、艾特金和罗西（Aitkin and Rocci, 2002）以及拉贝-赫斯基思等（Rabe-Hesketh et al., 2003a）也分析了在这种情形下的非参最大似然估计法。

本章说明了数学上非常相似的模型可以用于非常不同的问题中。例如，在指标和生存过程的联合模型中所用的模型结构与14.2节所讨论的饮食和心脏病的模型相似。如果去除指标模型中的随机截距，这两个应用模型则具有相同的结构，除了响应变量的类型不同以外（在指标和生存例子中为存续型，而在饮食和心脏病例子中为二项响应变量）。然而，对模型的解释却十分不同。

职业培训和子宫颈癌这两个例子的模型具有非常相似的结构。在这两个例子中，都针对子样本中观测到的二项协变量使用了一个潜类模型（服从以及暴露于疱疹病毒）。在这两个模型中，真实协变量的类别以及真实协变量与因变量的关系在两个子样本中都被假设为相同。在职业培训的例子中，潜类归属的进一步信息来自于关于服从的协变量，而在子宫颈癌例子中，其进一步的信息来自于两个子样本中关于疱疹病毒暴露的不齐全测量。

里拉德（Lillard, 1993）提供了一个关于多过程模型的非常有趣的应用例子，他提出了一个关于婚姻存续和生育的联立方程生存模型。目前，关于多过程和混合响应的潜变量模型，我们还仅仅处在理解其应用领域的初始阶段。然而，我们估计在这个领域将会有令人激动的进一步发展。

# 术语对照表

| 英文 | 中文 | 章节 |
| --- | --- | --- |
| absorbing event | 吸收事件 | Chp2 |
| accelerated failure time models | 加速失效时间模型 | Chp2 |
| adaptive Gaussian quadrature | 自适应高斯积分法 | |
| adaptive quadrature | 自适应积分法 | |
| additive effects | 叠加效应 | Chp2 |
| adjacent category logit models | 邻类对数概率模型 | Chp2 |
| adjusted empirical Bayes predictors | 修正经验贝叶斯预测量 | Chp6 |
| adjusted log ratio | 修正对数比值 | Chp11 |
| adjusted $R^2$ | 修正 $R^2$ | |
| aggregated data | 聚集数据 | Chp1 |
| analytical expression | 解析式 | |
| analytical integration | 解析积分 | |
| analytical solution/analytic solution | 解析解 | |
| antedependence | 前相关 | Chp3 |
| attenuated coefficient | 衰减系数 | Chp1 |
| autocorrelated errors | 自相关误差 | Chp3 |
| Autoregressive | 自回归的 | Chp3 |
| balanced mixed ANOVA models | 平衡混合 ANOVA 模型 | Chp6 |
| Bernoulli | 伯努利 | |
| beta binomial | 贝塔二项式 | Chp6 |
| beta distribution | 贝塔分布 | Chp2 |
| beta-binomial distribution | 贝塔-二项分布 | Chp2 |
| binary responses | 二项响应 | Chp2 |
| binomial logit-normal models(BLN) | 二项对数概率正态模型 | Chp1 |
| binomial logit-normal(BLN)model | 二项式对数概率-正态模型 | |
| biometrics | 生物计量学 | |
| biserial correlation | 双序列相关 | |
| bivariate normality | 二元正态性 | |
| bootstrap | 自助法 | |
| cartesian product quadrature | 笛卡儿乘积积分 | Chp6 |
| ceiling effect | 天花板效应 | Chp2 |
| censored response | 删失响应 | Chp2 |
| choice-based sampling | 择基抽样 | Chp2 |
| Cholesky decomposition | 乔莱斯基分解 | Chp6 |
| cluster randomized trial | 群组随机试验 | |
| clustered binary data | 丛聚二元数据 | Chp9 |

| | | |
|---|---|---|
| Cochran Q-test of homogeneity | 科克伦同质性 Q 检验 | Chp9 |
| cohort effect | 同期群效应 | Chp1 |
| common scale parameter | 公共尺度参数 | Chp2 |
| complementary log-log models | 余重对数模型 | Chp1 |
| complier average causal effect(CACE) | 服从者平均因果效应 | Chp14 |
| compound logistic-normal distribution | 复合对数概率正态分布 | Chp4 |
| concomitant variable | 伴随变量 | Chp4 |
| concurrent validity | 同期效度 | |
| conditional Gaussian autoregressive model | 条件高斯自回归模型 | |
| conditional likelihood | 条件似然 | Chp9 |
| conditional log-likelihood | 条件对数似然值 | |
| conditional maximum likelihood, CML | 条件最大似然 | |
| confidence limit | 置信限 | |
| confirmatory factor analysis | 验证性因子分析 | Chp3 |
| congeneric measurement model | 同属测量模型 | Chp3 |
| conjoint analysis | 结合分析 | Chp1 |
| *construct validity* | 构念效度 | |
| content validity | 内容效度 | |
| controlled clinical trials | 临床对照试验 | |
| convergent validity | 收敛效度 | Chp1 |
| Cook's distances | Cook 距离 | |
| Counterfactuals | 反事实性 | Chp1 |
| Cox regression | 考克斯回归 | Chp6 |
| credible intervals | 可信区间 | Chp9 |
| cross-sectional data | 截面数据 | Chp1 |
| cross-sectional survey | 截面调查 | |
| cross-sectional time-series data | 截面时间序列数据 | Chp3 |
| Crude Monte Carlo integration | 原始蒙特卡洛积分 | |
| cumulative hazard | 累积风险 | Chp2 |
| data augmentation algorithm | 数据增强算法 | |
| deletion residual | 删除残差 | Chp6,8 |
| delta method | 德尔塔方法 | |
| deviance information criterion(DIC) | 偏差信息标准 | Chp8 |
| diagnostic standard error | 诊断性标准误 | Chp6 |
| difference model | 差分模型 | Chp2 |
| differential item functioning(DIF) | 题器功能差异 | Chp9 |
| Differential measurement error | 差异性测量误差 | |
| Dirichlet process | 狄利克雷进程 | Chp4 |
| discrete-time | 离散时间 | Chp2 |
| discriminant validity | 判别效度 | |
| disease mapping | 疾病制图 | Chp1 |
| dispersion parameter | 离散参数 | Chp2 |
| divide-by-total models | 除总模型 | Chp2 |
| double expectation rule | 双重期望规则 | Chp4 |
| double-monotonicity | 双重单调性 | Chp3 |
| duration analysis | 存续分析 | Chp2 |

| | | |
|---|---|---|
| ecological fallacy | 生态学谬误 | Chp1 |
| eigenvalue | 本征值 | |
| EM algorithms | 期望最大化算法 | Chp1 |
| empirical identification | 经验辨识 | Chp5 |
| endogeneity | 内生性 | Chp3 |
| equivalent model | 等价模型 | Chp2 |
| error-component model | 误差成分模型 | Chp3 |
| exact solution | 精确解 | |
| exogeneity | 外生性 | Chp3 |
| expected cross-validation index(ECVI) | 期望交叉校验指标 | |
| exploded logit model | 激增对数概率模型 | Chp2 |
| exploratory factor analysis | 探索性因子分析 | Chp3 |
| exponential duration model | 指数存续模型 | Chp2 |
| factor loading | 因子负荷 | |
| factor score | 因子得分 | Chp7 |
| factorial invariance | 因子恒等性 | Chp3 |
| finite mixtures | 有限混合 | Chp3 |
| Fisher information | 费舍尔信息 | Chp6 |
| focussed information criterion(FIC) | 焦点信息标准 | Chp 8.4.3 |
| gamma distribution | 伽玛分布 | Chp2 |
| Gateaux derivative | 加托导数 | Chp6 |
| Gauss-Hermite quadrature | 高斯-埃尔米特积分 | Chp6 |
| Gaussian quadrature | 高斯积分 | Chp6 |
| general Markov model | 一般马尔可夫模型 | Chp9 |
| generalized estimating equations(GEE) | 广义估计方程 | Chp6 |
| generalized least squares | 广义最小二乘 | Chp3 |
| generalized MIMIC | 广义 MIMIC | Chp10 |
| generalized residuals | 广义残差 | Chp8 |
| Gibbs sampling | 吉布斯抽样 | Chp6 |
| graded response models | 等级响应模型 | Chp2 |
| gradient method | 梯度法 | |
| gradient vector | 梯度向量 | |
| GRC notation | GRC 标记法 | |
| growth curve model | 增长曲线模型 | |
| Gumbel distribution | 冈贝尔分布 | Chp1,13 |
| Hausman specification test | 豪斯曼设定检验 | Chp3 |
| heteroscedasticity | 异方差性 | Chp2 |
| hierarchical Bayesian model | 多层贝叶斯模型 | Chp6 |
| identity link | 同一性链接 | Chp6,9 |
| incremental fit index | 增量拟合指标 | |
| independence from irrelevant alternatives | 无关选项独立性 | Chp2 |
| interpolatory formulae | 插值公式 | Chp8 |
| intraclass correlation | 组内相关 | Chp3 |
| intrinsic autoregressive Gaussian(IGAR)model | 内在自回归高斯模型 | Chp11 |
| inverse gamma density | 逆伽玛密度 | Chp9 |
| inverse Gaussian | 逆高斯 | Chp2 |

| | | |
|---|---|---|
| irregular point | 非正则点 | |
| item response model | 题器响应模型 | |
| iterative generalized least squares(IGLS) | 迭代广义最小二乘法 | Chp6 |
| iteratively reweighted least squares(IRLS) | 迭代重加权最小二乘法 | Chp6 |
| Jackknife | 刀切法 | |
| Jacobian | 雅可比行列式 | |
| Jacobian matrix | 雅可比矩阵 | |
| joint likeihood | 联合似然 | Chp1 |
| joint maximum likelihood(JML) | 联合最大似然 | Chp6 |
| Lagrange multiplier | 拉格朗日乘数 | |
| Laplace approximation | 拉普拉斯近似 | |
| latent class models | 潜类模型 | Chp1 |
| latent profile models | 潜剖面模型 | Chp3 |
| latent scoring | 潜变量计分 | |
| latent traits | 潜特质 | |
| latent trajectory models | 潜轨迹模型 | Chp3 |
| latent variables | 潜变量 | Chp1 |
| Likelihood | 似然/似然性 | Chp1 |
| likelihood-ratio tests | 似然比值检验 | |
| linear by linear association model | 线性关联模型 | Chp2 |
| log odds | 对数比率 | |
| log ratio | 对数比值 | |
| Logistic | 对数概率 | |
| logit model | 对数概率模型 | Chp1 |
| logitstic distribution | 对数分布 | |
| log-linear model | 对数线性模型 | Chp3 |
| log-logistic duration model | 对数-对数概率存续模型 | Chp2 |
| log-normal duration model | 对数正态存续模型 | Chp2 |
| log-odds ratio | 对数优比 | |
| longitudinal analysis | 纵贯性分析 | Chp1 |
| manifest variable | 显变量 | |
| Markov Chain Monte Carlo(MCMC)method | 马尔可夫链蒙特卡洛法 | |
| Markov random fields | 马尔可夫随机场 | |
| maximum likelihood ratio(MLR) | 最大似然比值 | Chp6 |
| maximum simulated likelihood(MSL) | 最大模拟似然 | Chp6 |
| mean raw residual | 粗残差均值 | |
| mean squared error loss function | 损失函数均方差 | |
| mean total residual | 总残差均值 | |
| mean value theorem | 中值定理 | Chp6 |
| mean-centered | 均值中心化 | Chp3 |
| meta-analysis | 元分析 | Chp1 |
| method of simulated moments | 模拟动差估计法 | Chp1 |
| missing at random(MAR) | 随机性缺失 | |
| missing completely at random(MCAR) | 完全随机性缺失 | |
| mixed logit model | 混合对数概率模型 | |
| mixture regression model | 混合回归模型 | |

| | | |
|---|---|---|
| model specifications | 模型设定 | Chp1 |
| Monte Carlo Expectation Maximization(MCEM) | 蒙特卡洛期望最大化 | Chp6 |
| moving average residuals | 移动平均残差 | Chp3 |
| multidimensional scaling | 多维尺度分析 | Chp1 |
| multilevel proportional odds model | 多层次成比例比率模型 | |
| multimethod-multitrait design | 多方法-多特质设计 | Chp3 |
| multinomial probit model | 多项正态概率模型 | |
| multinomial response | 多项响应 | Chp2 |
| multiple membership model | 多重隶属模型 | Chp6 |
| multiple-indicator multiple-cause model | 多指标多原因模型(MIMIC) | Chp3 |
| multiplicative effects | 乘积效应 | Chp2 |
| multiplicative heteroscedasticity | 乘积异方差性 | Chp2 |
| multivariate normal marginal distributions | 多元正态边缘分布 | |
| multivariate quasi-likelihood approach | 多元准似然方法 | Chp3 |
| multivariate reliability | 多元信度 | Chp7 |
| negative binomial distribution | 负二项分布 | Chp2 |
| Newton-Raphson algorithm | 牛顿-拉夫逊算法 | |
| nomological validity | 法则效度 | Chp1 |
| nondifferential measurement error | 非差异性测量误差 | Chp1 |
| noninformative priors | 无信息先验 | Chp6,9 |
| nonlinear moment structures | 非线性矩结构 | |
| nonparametric maximum likelihood estimator(NPMLE) | 非参最大似然估计值 | Chp9 |
| Odds ratio | 优比 | Chp6 |
| one-sided censoring | 单侧删失 | Chp2 |
| one-way error component model | 单向误差成分模型 | Chp3 |
| one-way random effects model | 单向随机效应模型 | Chp3 |
| ordinal scaled probit factor model | 定序有衡 probit 因子模型 | Chp10 |
| ordinary least squares | 普通最小二乘法 | Chp3 |
| orthogonal matrix | 正交矩阵 | Chp3 |
| parallel measurement model | 平行测量模型 | Chp3 |
| parameter points | 参数点 | |
| partial credit model | 部分评分模型 | Chp2 |
| partial least squares | 偏最小二乘法 | Chp1 |
| penalized quasi-likelihood(PQL) | 罚伪似然 | |
| Poisson regression | 泊松回归 | Chp2 |
| polyserial correlation | 多序列相关 | Chp4 |
| positive definite weight matrix | 正定加权矩阵 | Chp6 |
| posterior mean squared error | 后验均方差 | |
| predictive validity | 预测效度 | |
| principal component analysis | 主成分分析 | Chp3 |
| prior information | 先验信息 | Chp3 |
| priori analysis | 先验分析 | Chp1 |
| probit model | 正态概率模型 | |
| profile likelihood | 剖面似然 | |
| profile log-likelihood | 剖面对数似然 | Chp9 |

| | | |
|---|---|---|
| proportional hazards model | 比例风险模型 | Chp14 |
| quasi Newton method | 准牛顿法 | Chp6 |
| quasi-likelihood method | 准似然方法 | Chp2 |
| quasi-Monte Carlo | 准蒙特卡洛 | Chp6 |
| Rao-Blackwellization | 劳-布莱克威尔法 | Chp6 |
| rating scale model | 等级尺度模型 | Chp3 |
| recursive model | 递归模型 | Chp3 |
| regression dilution | 回归稀释 | Chp1,3 |
| Residual correlation | 残相关 | |
| residual intraclass correlations | 组间残相关 | |
| residual standard deviation | 残标准差 | |
| residual variance | 残方差 | |
| root mean square error of approximation (RMSEA) | 近似误差均方根 | |
| sandwich estimator | 夹挤估计量 | Chp3 |
| scale parameter | 尺度参数 | Chp2 |
| semiparametric model | 半参数模型 | |
| singularity | 奇异性 | |
| small area estimation | 小域估计 | Chp1 |
| spherical quadrature | 球面积分 | Chp6 |
| Stereotype models | 铅版模型 | Chp2 |
| summed quadratic loss function | 加总二次损失函数 | Chp7 |
| survival analysis | 生存分析 | Chp1,2 |
| Taylor series | 泰勒级数 | |
| tetrachoric correlation | 四项相关 | Chp1,3 |
| tobit model | 托比模型 | Chp4 |
| truncated Hermite series | 截断埃尔米特级数 | |
| two-limit probit | 双重限制正态概率 | Chp2 |
| two-way error component mode | 双向误差成分模型 | Chp3 |
| unique factor | 独特因子 | Chp9 |
| validation sample | 校验样本 | Chp1,3 |
| variance components factor models | 方差成分因子模型 | Chp4 |
| variance components model | 方差成分模型 | Chp3 |
| Voxel | 三维像素 | |
| Wishart distribution | 威沙特分布 | Chp6 |
| zero-inflated binomial model | 零堆积二项式模型 | |
| zero-inflated Poisson model | 零堆积泊松模型 | Chp1 |

# 人名对照表

Akaike 赤池
Abramson 艾布拉姆森
Agresti 阿奎斯提
Aitchison 艾勤森
Aitkin 艾特金
Albert 阿尔伯特
Allenby 阿伦比
Allison 埃里森
Alwin 埃文
Anderson 安德森
Ansari 安萨里
Anscombe 安斯康姆(安斯康比)
Arbuckle 阿巴克尔
Arminger 阿明戈尔
Armstrong 阿姆斯特朗
Bahadur 巴哈杜尔
Bailey 拜利
Bagozzi 巴格兹
Banfield 班菲尔德
Barnes 巴恩斯
Bartholomew 巴托罗买
Bartlett 巴特利特
Bates 贝茨
Bechger 贝西格
Beggs 伯格斯
Bekker 贝克尔
Bennett 本尼特
Bensmail 本斯梅尔
Bente 本特
Bentler 本特勒
Berkhof 伯克霍夫
Berkson 伯克森
Berndt 博恩特
Best 贝斯特
Biggeri 比杰里
Blanchard 布兰查德
Bloxom 布洛克瑟姆
Bock 博科
Böckenholt 博肯哈特
Böhning 波宁
Bohrnstedt 伯恩斯特
Bollen 伯勒恩
Bonett 博奈特
Boomsma 波姆斯玛
Booth 布斯
Bosker 鲍斯克
Bottai 波太
Bottai 博塔伊
Box 巴克斯
Bradley 布拉德利
Brady 布拉迪
Breiman 布雷曼
Breslow 布里斯罗
Browne 布朗尼
Bryk 布莱克
Bunch 邦奇
Burnham 伯恩汉姆/伯纳姆
Busing 巴辛
Butler 巴特勒
Cameron 卡梅隆
Campbell 坎贝尔
Carey 凯里
Carmines 卡迈因斯
Carroll 卡洛尔
Cattell 卡特尔
Chapman 查普曼
Chechile 切奇利
Chesher 切舍尔
Chib 奇布
Cholesky 乔莱斯基
Christakis 克里斯塔基斯
Christensen 克里斯坦森
Christofferson 克里斯托弗森
Claeskens 克莱斯肯
Clayton 克雷顿
Clogg 克劳格

Cochran 科克伦
Congdon 康登
Converse 康弗斯
Cook 库克
Cormack 科马克
Coull 库尔
Cox 考克斯
Craig 克雷格
Cronbach 克朗巴哈克龙巴赫
Croon 克洛恩
Crouchley 克劳奇莱
Cudeck 科达克
Cuschwa 库西瓦
Dale 达勒
Danahy 德纳
Daniels 丹尼尔
Dave 戴夫
Davidson 达维森
Davidian 大卫迪安
Davies 戴维斯
Day 戴伊
Dayton 戴顿
DeBoeck 德伯克
DeSarbo 德萨伯
Deely 迪莱
Dempster 丹普斯特
Dey 戴伊
Diggle 迪歌
Dillon 迪伦
Donner 多纳
Draper 德雷珀
Dupacǒvá 杜帕科瓦
Durbin 杜宾
Druham 杜尔汉姆
Ecochard 埃克沙德
Edwards 爱德沃兹
Efron 艾弗朗
Einhorn 艾霍恩
Elrod 艾洛德
Engle 恩格
Everitt 艾弗瑞特
Ezzet 埃兹泽特
Fahrmeir 法梅尔
Farah 法拉
Firth 费斯
Fischer 费希尔
Fisher 费舍尔
Fitzmaurice 费兹马里斯
Fleiss 弗雷斯
Fokoué 福可
Follmann 弗曼
Formann 福尔曼
Fornell 弗奈尔
Fox 福克斯
Francis 弗朗西斯
Freedman 弗里德曼
Gallant 加兰特
Gart 加尔特
Gatsonis 盖特索尼斯
Gelman 葛尔曼
Gibbs 吉布斯
Gifford 吉福德
Gilks 吉尔克思
Glas 格拉斯
Goldberger 戈德伯格
Goldman 戈德曼
Goldstein 戈德斯坦恩
Goodman 古德曼
Gourieroux 古力洛克斯
Greene 格里尼
Grilli 格里利
Grove 格鲁夫
Gustafson 古斯塔夫松
Guttmann 古特曼
Haaijer 哈埃耶
Haberman 哈伯曼
Hagenaars 哈格纳斯
Hall 霍尔
Hambleton 汉布尔顿
Han 汉
Harper 哈珀
Harrell 哈瑞尔
Hartzel 哈泽尔
Harvey 哈维
Hauck 豪克
Hausman 豪斯曼
Heagerty 希格蒂
Heckman 赫克曼
Hedeker. 海德克
Heinan 埃南
Henly 亨莱
Heron 海伦
Hilden-Minton 海尔顿-明顿
Hildesheim 希尔德斯海姆

Hill 希尔
Hinkley 亨克莱
Hjort 约尔特
Hobert 霍伯特
Hodges 哈奇斯
Hogan 霍根
Hoijtink 霍伊廷克
Holford 霍尔福德
Horowitz 霍洛维茨
Hougaard 休加
Huber 哈伯尔
Huggins 哈金斯
Imbens 英本斯
Inglehart 英格尔哈特
Irish 艾里什
Jedidi 杰迪迪
Jenkins 詹金斯
Jennings 詹宁斯
Jo 乔
Jöreskog 杰拉斯科格
Kaldor 卡德
Kamakura 镰仓
Karim 卡里姆
Kaiser 凯瑟
Kass 卡斯
Katz 卡兹
Keane 齐内
Kelly 凯利
Kenkel 肯科
Kenward 肯沃德
Kestelman 科斯托曼
Kim 奇姆
King 金
Knott 诺特
Kohlmann 科洛曼
Krosnick 克罗斯尼克
Kuk 库克
Kupper 库珀
Küsters 库斯特
Laird 莱尔德
Lambert 兰伯特
Lange 朗日
Lasaffre 拉沙弗里
Lawley 劳利
Lawson 罗森
Lazarsfeld 拉扎斯菲尔德
Leamer 利莫尔
Lee 李
Lehmann 勒曼恩
Lenk 伦克
Leppik 莱皮克
Lerman 勒曼
Lesaffre 勒斯弗
Lewis 路易斯
Lieberman 李博曼
Lillard 里拉德
Lim 里姆
Lindley 林莱
Lindsay 林赛
Lindsey 林赛
Lipsitz 利普希茨
Listhaug 利斯塔乌戈
Little 雷托
Localio 罗加里奥
Long 朗
Longford 朗福特
Lord 劳德
Louis 路易斯
Luce 鲁斯
Luijben 鲁基本
Magidson 马吉德松
McCutcheon 麦克彻恩
MacCallum 麦克拉姆
MacKinnon 麦基侬
MacReady 麦克里迪
Maddala 马达拉
Magder 麦格德
Maggiotto 马吉奥多
Mallows 马罗斯
Maxwell 马克斯维尔
Manski 曼斯基
Marshall 马歇尔
Mason 梅森
Masters 马斯特斯
McCullagh 麦克莱
McCulloch 麦克拉奇
McDonald 麦当劳
McFadden 麦克法登
McGraw 麦克劳
McKelvey 麦凯维
McKendrick 麦肯迪克
Meehl 米赫尔
Meijer 迈尔
Meredith 梅雷迪斯

Miller 米勒
Mislevy 米斯勒维
Moffitt 墨菲特
Molenberghs 莫莱恩伯
Moustaki 穆斯塔奇
Mueller 穆勒
Mulaik 穆拉克
Muraki 穆拉齐
Muthén 穆森
Natarajan 纳塔拉杨
Naylor 奈勒
Nelder 奈尔德
Nelson 尼尔森
Neyman 奈曼
Novick 诺维卡
Nugent 纽金特
Nunnally 南奈利
Nychka 尼弛卡
Oakes 奥克斯
Olsson 奥尔森
Orchard 奥奇德
Ouwens 欧文斯
Patrick 帕德里克
Patterson 帕特森
Pawitan 帕维坦
Pearl 珀尔
Pearson 皮尔森
Pfeffermann 普费弗曼
Pickles 皮科尔斯
Pinheiro 皮涅罗
Pregibon 普莱吉本
Prentice 普伦迪斯
Rabe-Hesketh 拉贝-赫斯基思
Raftery 拉夫特里
Rampichini 拉姆皮奇尼
Rasbash 拉斯巴什
Rasch 拉什
Raudenbush 劳登布什
Revelt 利维特
Richardson 理查森
Richtand 里奇谭德
Rindskopf 林德斯科普夫
Rith 瑞斯
Robins 罗宾斯
Robinson 罗宾森
Rocci 罗西
Rochdale 罗奇代尔
Rodriquez 罗德里克
Roeder 罗德尔
Rosett 罗赛特
Rubin 鲁宾
Ryan 莱恩
Sackett 萨科特
Sammel 塞梅尔
Sampson 桑普森
Samejima 萨姆基玛
Sanathanan 萨纳特南
Sargent 萨金特
Satorra 萨托拉
Satten 沙特恩
Schafer 沙弗尔
Schilling 希林
Schluchter 希拉奇特
Schoenberg 肖恩伯格
Searle 希尔勒
Schall 夏尔
Shedden 谢登
Shrout 夏诺特
Silagy 希拉吉
Silvey 西尔维
Simar 希玛
Singer 辛格尔
Skaug 司考戈
Skrondal 斯科隆多
Snijders 斯奈德
Sommer 萨默尔
Spearman 斯皮尔曼
Spiegelhalter 斯皮格哈特
Spiegelman 斯皮格曼
Spiessens 斯皮森斯
Staelin 史德林
Steffey 斯特菲
Steiger 斯迭戈
Steinberg 斯滕伯格
Stern 斯特恩
Stevens 斯蒂文思
Stewart 斯特沃特
Stiratelli 斯特雷特里
Stott 斯托特
Strenio 斯特尼奥
Suppes 萨普斯
Swaminathan 斯瓦米纳坦
Takane 高岭
Tanner 唐纳

# 参考文献


Aalen, O. O. 1988. Heterogeneity in survival analysis. *Statistics in Medicine*, 7, 1121-1137.

Abramson, P. R. 1983. *Political Attitudes in America*. San Francisco, CA: Freeman.

Adams, R. C. 1975. On the use of a marker variable in lieu of its factor-analytic construct. *Perceptual and Motor Skills*, 41, 665-666.

Agresti, A. 1994. Simple capture-recapture models permitting unequal catchability and variable sampling effort. *Biometrics*, 50,494-500.

Agresti, A. 1996. *Introduction to Categorical Data Analysis*. New York: Wiley.

Agresti, A. 2002. *Categorical Data Analysis (Second Edition)*. Hoboken, NJ: Wiley.

Agresti, A., & Hartzel, J. 2000. Tutorial in biostatistics. Strategies for comparing treatments on a binary response with multi-centre data. *Statistics in Medicine*, 19, 1115-1139.

Agresti, A., & Natarajan, R. 2001. Modeling clustered ordered categorical data: A survey. *International Statistical Review*, 69, 345-371.

Aish, A.-M., & Jöreskog, K. G. 1989. *Political Efficacy: Measurement and Dimensionality*. Tech. rept. 89-4. Department of Statistics, Uppsala University.

Aitchison, J., & Silvey, S. 1957. The generalization of probit analysis to the case of multiple responses. *Biometrika*, 44, 131-140.

Aitkin, M. 1996. A general maximum likelihood analysis of overdispersion in generalised linear models. *Statistics and Computing*, 6, 251-262.

Aitkin, M. 1999a. A general maximum likelihood analysis of variance components in generalized linear models. *Biometrics*, 55, 117-128.

Aitkin, M. 1999b. Meta-analysis by random effect modelling in generalised linear models. *Statistics in Medicine*, 18, 2343-2351.

Aitkin, M., & Longford, N. T. 1986. Statistical modeling issues in school effectiveness studies. *Journal of the Royal Statistical Society, Series A*, 149, 1-43.

Aitkin, M., & Rocci, R. 2002. A general maximum likelihood analysis of measurement error in generalised linear models. *Statistics and Computing*, 12, 163-174.

Aitkin, M., & Rubin, D. B. 1985. Estimation and hypothesis testing in finite mixture distributions. *Journal of the Royal Statistical Society, Series B*, 47, 67-75.

Aitkin, M., Anderson, D., & Hinde, J. 1981. Statistical modeling of data on teaching styles. *Journal of the Royal Statistical Society, Series A*, 144,419-461.

Aitkin, M., Anderson, D., Francis, B. J., & Hinde, J. 1989. *Statistical Modelling in GLIM*. Oxford: Oxford University Press.

Aitkin, M., Anderson, D., Francis, B. J., & Hinde, J. 2004. *Statistical Modelling in GLIM (Second Edition)*. Oxford: Oxford University Press.

Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. *Pages* 267-281 *of*: Petrov, B. N., & Csaki, F. (eds), *Second International Symposium on Information Theory*. Budapest: Academiai Kiado.

Akaike, H. 1987. Factor analysis and AIC. *Psychometrika*, 52, 317-332.

Albert, J. H. 1992. Bayesian estimation of normal ogive item response curves using Gibbs sampling. *Journal of Educational Statistics*, 17, 251-269.

Albert, J. H., & Chib, S. 1993. Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 88, 669-679.

Albert, J. H., & Chib, S. 1995. Bayesian residual analysis for binary response regression models. *Biometrika*, 82,747-759.

Albert, J. H., & Chib, S. 1997. Bayesian tests and model diagnostics in conditionally independent hierarchical models. *Journal of the American Statistical Association*, 92, 916-925.

Albert, P. S. 1999. Tutorial in biostatistics. Longitudinal data analysis (repeated measures) in clinical trials. *Statistics in Medicine*, 18, 1707-1732.

Albert, P. S., & Follmann, D. A. 2000. Modeling repeated count data subject to informative dropout. *Biometrics*, 56, 667-677.

Allenby, G. M., & Lenk, P. J. 1994. Modeling household purchase behavior with logistic normal regression. *Journal of the*

*American Statistical Association*, 89, 1218-1231.

Allenby, G. M., Arora, N., & Ginter, J . L. 1998. On the heterogeneity of demand. *Journal of Marketing Research*, 35, 384-389.

Allison, P. D. 1984. *Event History Analysis. Regression for Longitudinal Event Data.* Sage University Paper Series on Quantitative Applications in the Social Sciences. Newbury Park, CA: Sage.

Allison, P. D., & Christakis, N. A. 1994. Logit models for sets of ranked items. *Pages* 199-228 *of*: Marsden, P. V. (ed), *Sociological Methodology* 1994. Oxford: Blackwell.

Alwin, D. F. 1989. Problems in the estimation and interpretation of the reliability of survey data. *Quality and Quantity*, 23, 277-331.

Alwin, D. F., & Krosnick, J. A. 1985. The measurement of values in surveys: A comparison of ratings and rankings. *Public Opinion Quarterly*, 49, 535-552.

American Psychological Association (APA), American Educational Research Association (AERA), & National Council on Measurement in Education (NCME). 1974. *Standards for Educational and Psychological Tests.* Washington, DC: American Psychological Association.

Andersen, E. B. 1970. Asymptotic properties of conditional maxinmm likelihood estimators. *Journal of the Royal Statistical Society, Series B*, 32,283-301.

Andersen, E. B. 1973. *Conditional Inference and Models for Measuring.* Copenhagen: Mentalhygiejnisk Forsknings Institut.

Andersen, E. B. 1980. *Discrete Statistical Models with Social Science Applications.* Amsterdam: North-Holland.

Andersen, P. K., & Gill, R. D. 1982. Cox's regression model for counting processes: A large sample study. *The Annals of Statistics*, 10, 1100-1120.

Andersen, P. K., Borgan, Ø., Gill, R. D., & Keiding, N. 1993. *Statistical Models Based on Counting Processes.* New York: Springer.

Anderson, J. A. 1984. Regression and ordered categorical variables. *Journal of the Royal Statistical Society, Series B*, 46, 1-30.

Anderson, T. W. 1963. The use of factor analysis in the statistical analysis of multiple time series. *Psychometrika*, 28, 1-25.

Anderson, T. W. 2003. *Introduction to Multivariate Statistical Analysis (Third Edition).* New York: Wiley.

Anderson, T. W., & Rubin, H. 1956. Statistical inference in factor analysis. *Pages* 111-150 *of*: Neyman, J. (ed), *Proceedings of the Third Berkeley Symposium On Mathematical Statistics and Probability.* Berkeley: University of California Press.

Andrews, F. M., & Withey, S. B. 1976. *Social Indicators of Well-Being: Americans' Perceptions of Life Quality.* New York: Plenum.

Andrich, D. 1978. A rating formulation for ordered response categories. *Psychometrika*, 43, 561-573.

Angrist, J. D., Imbens, G. W., & Rubin, D. B. 1996. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91,444-455.

Ansari, A., & Jedidi, K. 2000. Bayesian factor analysis for multilevel binary observations. *Psychometrika*, 65, 475-496.

Appelbaum, M. I., & McCall, R. B. 1983. Design and analysis in developmental psychology. *Pages* 415-476 *of*: Mussen, P. H. (ed), *Handbook of Child Psychology*, vol. I. New York: Wiley.

Arbuckle, J. L., & Nugent, J. H. 1973. A general procedure for parameter estimation for the law of comparative judgment. *British Journal of Mathematical and Statistical Psychology*, 26, 240-260.

Arbuckle, J. L., & Wothke, W. 1999. *Amos* 4.0 *Users' Guide.* Chicago, IL: SmallWaters Corporation.

Arbuckle, J. L., & Wothke, W. 2003. *Amos* 5.0 *Update to the Amos User's Guide.* Chicago, IL: SmallWaters Corporation.

Arminger, G., & Küsters, U. 1988. Latent trait models with indicators of mixed measurement levels. *Pages* 51-73 *of*: Langeheine, R., & Rost, J. (eds), *Latent Trait and Latent Class Models.* New York: Plenum.

Arminger, G., & Küsters, U. 1989. Construction principles for latent trait models. *Pages* 369-393 *of*: Clogg, C. C. (ed), *Sociological Methodology* 1989. Oxford: Blackwell.

Arminger, G., & Muthén, B. O. 1998. A Bayesian approach to nonlinear latent variable models using the Gibbs sampler and the Metropolis-Hastings algorithm. *Psychometrika*, 63, 271-300.

Arminger, G., Wittenberg, J., & Schepers, A. 1996. *MECOSA* 3 *User Guide.* Friedrichsdorf: ADDITIVE.

Arminger, G., Stein, P., & Wittenberg, J. 1999. Mixtures of conditional mean- and covariance-structure models. *Psychometrika*, 64, 475-494.

Armstrong, J. S. 1967. Derivation of theory by means of factor analysis or Tom Swift and his electric factor analysis machine. *The American Statistician*, 21, 17-21.

Bagozzi, R. P. 1980. *Causal Models in Marketing.* New York: Wiley.

Bagozzi, R. P. 1981. An examination of the validity of two models of attitude. *Multivariate Behavioral Research*, 16, 323-359.

Bahadur, R. R. 1961. A representation of the joint distribution of responses to $n$ dichotomous items. *Pages* 158-168 *of*: Solomon, H. (ed), *Studies in Item Analysis and Prediction.* Stanford Mathematical Studies in the Social Sciences, vol. VI. Stanford, CA: Stanford University Press.

Bailey, K. R. 1987. Inter-study differences: How should they influence the interpretation and analysis of results? *Statistics in Medicine*, 6, 351-358.

Baltagi, B. 2001. *The Econometrics of Panel Data (Second Edition).* London: Wiley.

Banfield, J. D., & Raftery, A. E. 1993. Model-based Gaussian and non-Gaussian clustering. *Biometrics*, 49, 803-821.

Barndorff-Nielsen, O. E. 1978. *Information and Exponential Families in Statistical Theory*. New York: Wiley.

Barnes, S. H., Kaase, M., Allerbeck, K. R., Farah, B. J., Heunks, F., Inglehart, R., Jennings, M. K., Klingemann, H. D., Marsh, H., & Rosenmayr, L. 1979. *Political Action. Mass Participation in Five Western Democracies*. Beverly Hills, CA: Sage.

Barnett, V., & Lewis, T. 1984. *Outliers in Statistical Data*. London: Wiley.

Baron, R. M., & Kenny, D. A. 1986. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology*, 51, 1173-1182.

Bartholomew, D. J. 1981. Posterior analysis of the factor model. *British Journal of Mathematical and Statistical Psychology*, 34, 93-99.

Bartholomew, D. J. 1987. *Latent Variable Models and Factor Analysis*. Oxford: Oxford University Press.

Bartholomew, D. J. 1988. The sensitivity of latent trait analysis to the choice of prior distribution. *British Journal of Mathermatical and Statistical Psychology*, 41, 101-107.

Bartholomew, D. J. 1991. Latent variable methods. *In*: Lovie, P., & Lovie, A. D. (eds), *New developments in statistics for psychology and the social sciences*, vol. 2. London: BPS Books and Routledge.

Bartholomew, D. J. 1993. Estimating relationships between latent variables. *Sankhya*, 55, 409-419.

Bartholomew, D. J. 1994. Bayes' theorem in latent variable modelling. *Pages* 41-50 *of*: Freeman, P. R., & Smith, A. F. M. (eds), *Aspects of uncertainty: A tribute to D. V. Lindley*. London: Wiley.

Bartholomew, D. J., & Knott, M. 1999. *Latent Variable Models and Factor Analysis*. London: Arnold.

Bartholomew, D. J., & Leung, S. O. 2002. A goodness of fit test for sparse $2^p$ contingency tables. *British Journal of Mathematical and Statistical Psychology*, 55, 1-15.

Bartlett, M. S. 1937. The statistical conception of mental factors. *British Journal of Psychology*, 28, 97-104.

Bartlett, M. S. 1938. Methods of estimating mental factors. *Nature*, 141, 609-610.

Bayarri, M. J., & Berger, J. O. 2000. P-values for composite null models. *Journal of the American Statistical Association*, 95, 1127-1142.

Bechger, T. M., Verhelst, N. D., & Verstralen, H. H. F. M. 2001. Identifiability of nonlinear logistic test models. *Psychometrika*, 66, 357-372.

Bechger, T. M., Verstralen, H. H. F. M., & Verhelst, N. D. 2002. Equivalent linear logistic test models. *Psychometrika*, 67, 123-136.

Becker, M. P., Yang, I., & Lange, K. 1997. EM algorithms without missing data. *Statistical Methods in Medical Research*, 6, 37-53.

Beggs, S., Cardell, S., & Hausman, J. A. 1981. Assessing the potential demand for electric cars. *Journal of Econometrics*, 16, 1-19.

Bekker, P. A., Merckens, A., & Wansbeek, T. J. 1994. *Identification, Equivalent Models and Computer Algebra*. Boston: Academic Press.

Belsley, D. A., Kuh, E., & Welsch, R. E. 1980. *Regression Diagnostics*. London: Wiley.

Bennett, S. 1983. Analysis of survival data by the proportional odds model. *Statistics in Medicine*, 2, 273-277.

Bensmail, H., Celeux, G., Raftery, A. E., & Robert, C. P. 1997. Inference in model-based cluster analysis. *Statistics and Computing*, 7, 1-10.

Bentler, P. M. 1978. The interdependence of theory, methodology and empirical data. *In*: Kandel, D. (ed), *Longitudinal Research On Drug Use. Empirical Findings And Methodological Issues*. Washington: Hemisphere.

Bentler, P. M. 1980. Multivariate analysis with latent variables: Causal Modelling. *Annual Review of Psychology*, 31, 419-456.

Bentler, P. M. 1982. Linear systems with multiple levels and types of latent variables. *Pages* 101-130 *of*: Jöreskog, K. G., & Wold, H. (eds), *Systems Under Indirect Observation-Causality ★ Structure ★ Prediction-Part I*. Amsterdam: North-Holland.

Bentler, P. M. 1995. *EQS Structural Equation Program Manual*. Encino, CA: Multivariate Software.

Bentler, P. M., & Bonett, D. G. 1980. Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88, 588-606.

Bergan, J. R. 1988. Latent variable techniques for measuring development. *Pages* 233-261 *of*: Langeheine, R., & Rost, J. (eds), *Latent Class and Latent Trait Models*. New York: Plenum Press.

Berger, J. O., & Delampady, M. 1987. Testing precise hypotheses (with discussion). *Statistical Science*, 2, 317-352.

Berger, J. O., & Sellke, T. 1987. Testing a point null hypothesis: The irreconcilability of $p$ values and evidence (with discussion). *Journal of the American Statistical Association*, 82, 112-139.

Berkey, C. S., Hoaglin, D. C., Mosteller, F., & Colditz, G. A. 1995. Random effects regression model for meta-analysis. *Statistics in Medicine*, 14, 394-411.

Berkhof, J., & Snijders, T. A. B. 2001. Variance component testing in multilevel models. *Journal of Educational and Behavioral Statistics*, 26, 132-152.

Berkson, J. 1950. Are there two regressions? *Journal of the American Statistical Association*, 45, 164-80.

Bernardinelli, L., & Montomoli, C. 1992. Empirical Bayes versus fully Bayesian analysis of geographical variation in disease risk. *Statistics in Medicine*, 11, 983-1007.

Berndt, E. R., Hall, B. H., Hall, R. E., & Hausman, J. A. 1974. Estimation and inference in nonlinear structural models. *Annals of Economic and Social Measurement*, 3,653-666.

Besag, J. 1974. Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society, Series B*, 36, 192-236.

Besag, J. 1986. On the statistical analysis of dirty pictures (with discussion). *Journal of the Royal Statistical Society, Series B*, 48, 259-302.

Besag, J., York, J., & Mollie, A. 1991. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43, 1-21.

Biggeri, A., Marchi, M., Lagazio, C., Martuzzi, M., & Böhning, D. 2000. Non-parametric maximum likelihood estimators for disease mapping. *Statistics in Medicine*, 19, 2539-2554.

Binder, D. A. 1983. On the variances of asymptotically normal estimators from complex surveys. *International Statistical Review*, 51,279-292.

Birnbaum, A. 1968. Test scores, sufficient statistics, and the information structures of tests. *Pages* 425-435 *of*: Lord, F. M., & Noviek, M. R. (eds), *Statistical Theories of Mental Test Scores*. Reading: Addison-Wesley.

Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W. 1975. *Discrete Multivariate Analysis. Theory and Practice*. Cambridge: MIT Press.

Blåfield, E. 1980. *Clustering of Observation from Finite Mixtures with Structural Information*. Ph. D. thesis, Jyväskyla studies in computer science, economics, and statistics, Jyväskyla.

Blalock, H. M. 1960. *Social Statistics*. New York: McGraw-Hill.

Bliss, C. I. 1934. The method of probits. *Science*, 79, 38-39.

Bloxom, B. 1972. The simplex in pair comparisons. *Psychometrika*, 37, 119-136.

Bock, R. D. 1972. Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, 37, 29-51.

Bock, R. D. 1983. The discrete Bayesian. *Pages* 103-115 *of*: Wainer, H., & Messick, S. (eds), *Principals of Modern Psychological Measurement*. Hillsdale, NJ: Erlbaum.

Bock, R. D. 1985. Contributions of empirical Bayes and marginal maximum likelihood methods to the measurement of individual differences. *Pages* 75-99 *of*: Roskam, E. E (ed), *Measurement and Personality Assessment*. Amsterdam: Elsevier.

Bock, R. D., & Aitkin, M. 1981. Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, 46, 443-59.

Bock, R. D., & Lieberman, M. 1970. Fitting a response model for *N* dichotomously scored items. *Psychometrika*, 35, 179-197.

Bock, R. D., & Mislevy, R. J. 1982. Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement*, 6, 431-444.

Bock, R. D., & Schilling, S. G. 1997. High-dimensional full-information item factor analysis. *Pages* 164-176 *of*: Berkane, M. (ed), *Latent Variable Modelling and Applications to Causality*. New York: Springer.

Bock, R. D., Gibbons, R. D., Schilling, S. G., Muraki, E., Wilson, D. T., & Wood, R. 1999. *TESTFACT: Test Scoring, Item Statistics, and Full Information Item Factor Analysis*. Chicago, IL: Scientific Software International.

Böckenholt, U. 1992. Thurstonian representation for partial ranking data. *British Journal of Mathematical and Statistical Psychology*, 45, 31-39.

Böckenholt, U. 1993. Application of Thurstonian models to ranking data. *In*: Fligner, M. A., & Verducci, J. S. (eds), *Probability Models and Statistical Analysis for Ranking Data*. New York: Springer.

Böckenholt, U. 2001a. Hierarchical modeling of paired comparison data. *Psychological Methods*, 6, 49-66.

Böckenholt, U. 2001b. Mixed-effects analyses of rank-ordered data. *Psychometrika*, 66, 45-62.

Böhning, D. 1982. Convergence of Simar's algorithm for finding the maximum likelihood estimate of a compound Poisson process. *The Annals of Statistics*, 10, 1006-1008.

Böhning, D. 2000. *Computer-Assisted Analysis of Mixtures and Applications. Meta-Analysis, Disease Mapping and Others*. London: Chapman & Hall.

Böhning, D., Ekkehart, D., Schlattmann, P., Mendonça, L., & Kircher, U. 1999. The zeroinflated Poisson model and the decayed, missing and filled teeth index in dental epidemiology. *Journal of the Royal Statistical Society, Series A*, 162, 195-209.

Bohrnstedt, G. W. 1983. Measurement. *Pages* 69-121 *of*: Rossi, P. H., Wright, J. D., & Anderson, A. B. (eds), *Handbook of Survey Research*. New York: Academic.

Bollen, K. A. 1989. *Structural Equations with Latent Variables*. New York: Wiley.

Bollen, K. A. 2002. Latent variables in psychology and the social sciences. *Annual Review of Psychology*, 53, 605-634.

Bollen, K. A., & Arminger, G. 1991. Observational residuals in factor analysis and structural equation models. *Pages* 235-262 *of*: Marsden, P. V. (ed), *Sociological Methodology* 1991. Oxford: Blackwell.

Bollen, K. A., & Long, J. S. (eds). 1993. *Testing Structural Equation Models*. Newbury Park, CA: Sage.

Booth, J. G., & Hobert, J. P. 1999. Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. *Journal of the Royal Statistical Society, Series B*, 62, 265-285.

Booth, J. G., Hobert, J. P., & Jank, W. 2001. A survey of Monte Carlo algorithms for maximizing the likelihood of a two-

stage hierarchical model. *Statistical Modelling*, 1, 333-349.

Bottai, M. 2003. Confidence regions when the Fisher information is zero. *Biometrika*, 90, 73-84.

Box, G. E. P., Hunter, W. G., & Hunter, J. S. 1978. *Statistics for Experimenters.* New York: Wiley.

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. 1994. *Time Series Analysis: Forecasting and Control.* Englewood Cliffs, NJ: Prentice-Hall.

Bozdogan, H. 1987. Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, 52, 345-370.

Bradley, R. E., & Terry, M. E. 1952. Rank analysis of incomplete block designs I. The method of paired comparisons. *Biometrika*, 39, 324-345.

Brady, H. E. 1989. Factor and ideal point analysis for interpersonally incomparable data. *Psychometrika*, 54, 181-202.

Brady, H. E. 1990. Dimensional analysis of ranking data. *American Journal of Political Science*, 34, 1017-1048.

Breckler, S. 1990. Applications of covariance structure modeling in psychology: Cause for concern? *Psychological Bulletin*, 107, 260-273.

Breiman, L. 2001. Statistical modeling: The two cultures. *Statistical Science*, 16, 199-215.

Brennan, R. L. 2001. *Generalizability Theory.* New York: Springer.

Breslow, N. E. 1974. Covariance analysis of censored survival data. *Biometrics*, 30, 89-100.

Breslow, N. E. 2003. Whither PQL? *UW Biostatistios Working Paper Series*, 1-25. Downloadable from http://www.bepress.com/uwbiostat/paper192/.

Breslow, N. E., & Clayton, D. G. 1993. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88, 9-25.

Breslow, N. E., & Day, N. 1980. *Statistical Methods in Cancer Research. Vol I-The Analysis of Case-Control Studies.* Lyon: IARC.

Breslow, N. E., & Day, N. 1987. *Statistical Methods in Cancer Research. Vol II-The Design and Analysis of Cohort Studies.* Lyon: IARC.

Breslow, N. E., & Lin, X. 1995. Bias correction in generalised linear mixed models with a single component of dispersion. *Biometrika*, 82, 81-91.

Breslow, N. E., Leroux, B., & Platt, R. 1998. Approximate hierarchical modeling of discrete data in epidemiology. *Statistical Methods in Medical Research*, 4, 49-62.

Brinch, C. 2001. *Non-parametric Identification of Mixed Hazards Models.* Ph. D. thesis, Department of Economics, University of Oslo.

Brody, R. A., & Page, B. I. 1972. Comment: The assessment of policy voting. *American Political Science Review*, 66, 450-458.

Browne, M. W. 1984. Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, 37, 62-83.

Browne, M. W. 2000. Cross-validation methods. *Journal of Mathematical Psychology*, 44, 108-132.

Browne, M. W., & Arminger, G. 1995. Specification and estimation of mean- and covariance-structure models. *Pages* 185-249 *of*: Arminger, G., Clogg, C. C., & Sobel, M. E. (eds), *Handbook of Statistical Modelling for the Social and Behavioral Sciences.* New York: Plenum Press.

Browne, M. W., & Cudeck, R. 1989. Single sample cross-validation indices for covariance structures. *Multivariate Behavioral Research*, 24, 445-455.

Browne, M. W., & Cudeck, R. 1993. Alternative ways of assessing model fit. *Pages* 136-162 *of*: Bollen, K. A., & Long, J. S. (eds), *Testing Structural Equation Models.* Newbury Park, CA: Sage.

Browne, W. J. 1998. *Applying MCMC Methods to Multi-Level Models.* Ph. D. thesis, Department of Statistics, University of Bath.

Browne, W. J. 2004. *MCMC estimation in MLwiN.* London: Instiute of Education. Downloadable from http://multilevel.ioe.ac.uk/download/manuals.html.

Browne, W. J., & Draper, D. 2004. A comparison of Bayesian and likelihood methods for fitting multilevel models. *Submitted.* Downloadable from http://multilevel.ioe.ac.uk/team/materials/wbrssa.pdf.

Bunch, D. S. 1991. Estimability in the multinomial probit model. *Transportation Research B*, 25, 1-12.

Burnham, K. P., & Anderson, D. R. 2002. *Model Selection and Multimodel Inference. A Practical Information-Theoretic Approach (Second Edition).* New York: Springer.

Buse, A. 1982. The Likelihood Ratio, Wald, and Lagrange Multiplier tests: An expository note. *The American Statistician*, 36, 153-157.

Busemeyer, J. R., & Jones, L. E. 1983. Analysis of multiplicative combination rules when the causal variables are measured with error. *Psychological Bulletin*, 93, 549-562.

Busing, F. M. T. A. 1993. *Distribution characteristics of variance estimates in two-level models; A Monte Carlo study. Preprint PRM* 93-04. Tech. rept. Department of Psychometrics and Research Methodology, Leiden University, Leiden, The Netherlands.

Busing, F. M. T. A., Meijer, E., & van der Leden, R. 1994. *MLA. Software for multilevel analysis of data with two levels. Users guide for version* 1.0*b*. *Preprint PRM* 94-01. Tech. rept. Department of Psychometrics and Research Methodology, Leiden University, Leiden, The Netherlands.

Butler, J. S., & Moffitt, R. 1982. A computationally efficient quadrature procedure for the one-factor multinomial probit model. *Econometrica*, 50, 761-764.

Butler, S. M., & Louis, T. A. 1992. Random effects models with nonparametric priors. *Statistics in Medicine*, 11, 1981-2000.

Cameron, A. C., & Trivedi, P. K. 1998. *Regression Analysis of Count Data*. Cambridge: Cambridge University Press.

Campbell A., Gurin, G., & Miller, W. E. 1954. *The Voter Decides*. Evanston IL: Row Peterson.

Campbell, A., Converse, P. E., & Rodgers, W. L. 1976. *The Quality of American Life*. New York, NY: Sage.

Cappellari, L., & Jenkins, S. P. 2003. Mnltivariate probit regression using simulated maximum likelihood. *The Stata Journal*, 3, 278-294.

Carey, V. C., Zeger, S. L., & Diggle, P. J. 1993. Modelling multivariate binary data with alternating logistic regressions. *Biometrika*, 80, 517-526.

Carlin, B. P., & Louis, T. A. 2000. *Bayes and Empirical Bayes Methods for Data Analysis (Second Edition)*. Boca Raton: Chapman & Hall/CRC.

Carlin, J. B. 1992. Meta-analysis for 2 ×2 tables: A Bayesian approach. *Statistics in Medicine*, 11, 141-158.

Carpenter, J., Goldstein, H., & Rasbash, J. 1999. A non-parametric bootstrap for multilevel models. *Multilevel Modelling Newsletter*, 11, 2-5.

Carroll, J. D., & Chang, J. J. 1972. *IDIOSCAL (Individual differences in orientation SCALing): Generalization of INDSCAL allowing IDIOsyncratic reference systems as well as analytic approximation to INDSCAL.* Paper presented at the meeting of the Psychometric Society.

Carroll, R. J., Gail, M. H., & Lubin, J. H. 1993. Case-control studies with errors in covariates. *Journal of the American Statistical Association*, 88, 185-199.

Carroll, R. J., Ruppert, D., & Stefanski, L. A. 1995a. *Measurement Error in Nonlinear Models.* London: Chapman & Hall.

Carroll, R. J., Wang, S., & Wang, C. Y. 1995b. Prospective analysis of logistic case-control studies. *Journal of the American Statistical Association*, 90, 157-169.

Carroll, R. J., Roeder, K., & Wasserman, L. 1999. Flexible parametric measurement error models. *Biometrics*, 86, 44-54.

Casella, G., & George, E. I. 1992. Explaining the Gibbs sampler. *The American Statistician*, 46, 167-174.

Cattell, R. B. 1949. The dimensions of culture patterns by factorization of national characters. *Journal of Abnormal and Social Psychology*, 44, 443-469.

Cattell, R. B. 1966. The scree test for the number of factors. *Multivariate Behavioral Research*, 1,245-276.

Chalmers, I. 1993. The Cochrane Collaboration: Preparing, maintaining, and disseminating systematic reviews of the effects of health care. *Annals of the New York Academy of Science*, 703, 156-165.

Chamberlain, G. 1980. Analysis of covariance with qualitative data. *Review of Economic Studies*, 47, 225-238.

Chan, W., & Bentler, P. M. 1998. Covariance structure analysis of ordinal ipsative data. *Psychometrika*, 63, 369-399.

Chao, A., Tsay, P. K., Lin, S.-H., Shau, W.-Y., & Chao, D.-U. 2001. Tutorial in biostatistics. The applications of capture-recapture models to epidemiological data. *Statistics in Medicine*, 20, 3123-3157.

Chapman, R. G., & Staelin, R. 1982. Exploiting rank ordered choice set data within the stochastic utility model. *Journal of Marketing Research*, 14, 288-301.

Chatterjee, S., & Hadi, A. S. 1988. *Sensitivity Analysis in Linear Regression.* London: Wiley.

Chechile, R. 1977. Likelihood and posterior identification: Implications for mathematical psychology. *British Journal of Mathematical and Statistical Psychology*, 30, 177-184.

Chen, C.-F. 1981. The EM approach to the multiple indicators and multiple causes model via the estimation of the latent variable. *Journal of the American Statistical Association*, 76, 704-708.

Chesher, A., & Irish, M. 1987. Residual analysis in the grouped and censored linear model. *Journal of Econometrics*, 34, 33-61.

Chib, S., & Greenberg, E. 1995. Understanding the Metropolis-Hastings algorithm. *The American Statistician*, 49, 327-335.

Chib, S., & Hamilton, B. H. 2002. Semiparametric Bayes analysis of longitudinal data treatment models. *Journal of Econometrics*, 110, 67-89.

Christensen, E., Schlichting, P., Andersen, P. K., Fauerholdt, L., Schou, G., Pedersen, B. V., Juhl, E., Poulsen, H., & Thygstrup, N. 1986. Updating prognosis and therapeutic effect evaluation in cirrhosis using Cox's multiple regression model for time dependent variables. *Scandinavian Journal of Gastroenterology*, 21, 163-174.

Christofferson, A. 1975. Factor analysis of dichotomized variables. *Psychometrika*, 40, 5-32.

Claeskens, G., & Hjort, N. L. 2003. The focussed information criterion (with discussion). *Journal of the American Statistical Association*, 98, 900-916.

Clarkson, D. B., & Zhan, Y. H. 2002. Using spherical-radial quadrature to fit generalized linear mixed effects models. *Journal of Computational and Graphical Statistics*, 11, 639-659.

Clayton, D. G. 1978. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65, 141-151.

Clayton, D. G. 1988. The analysis of event history data: A review of progress and outstanding problems. *Statistics in Medicine*, 7, 819-841.

Clayton, D. G. 1992. Models for the analysis of cohort and case-control studies with inaccurately measured exposures. *Pages* 301-331 *of*: Dwyer, J. H., Feinlieb, M., Lippert, P., & Hoffmeister, H. (eds), *Statistical Models for Longitudinal Studies*

*on Health*. New York: Oxford University Press.

Clayton, D. G. 1996a. Contribution to the discussion of a paper by Y. Lee and J. A. Nelder. *Journal of the Royal Statistical Society, Series B*, 58, 657-659.

Clayton, D. G. 1996b. Generalized linear mixed models. *In*: Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (eds), *Markov Chain Monte Carlo in Practice*. London: Chapman & Hall.

Clayton, D. G., & Cuzick, J. 1985. Multivariate generalizations of the proportional hazards model. *Journal of the Royal Statistical Society, Series A*, 148, 82-108.

Clayton, D. G., & Hills, M. 1993. *Statistical Models in Epidemiology*. Oxford: Oxford University Press.

Clayton, D. G., & Kaldor, J. 1987. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics*, 43, 671-681.

Clayton, D. G., & Rasbash, J. 1999. Estimation in large crossed random-effect models by data augmentation. *Journal of the Royal Statistical Society, Series A*, 162,425-436.

Cliff, N. 1983. Some cautions concerning the application of causal modeling methods. *Multivariate Behavioral Research*, 18, 115-126.

Clogg, C. C. 1979. Some latent structure models for the analysis of Likert-type data. *Social Science Research*, 8, 287-301.

Clogg, C. C. 1988. Latent class models for measuring. *Pages* 173-205 *of*: Langeheine, R., & Rost, J. (eds), *Latent Trait and Latent Class Models*. New York: Plenum Press.

Clogg, C. C. 1995. Latent class models. *Pages* 311-359 *of*: Arminger, G., Clogg, C. C., & Sobel, M. E. (eds), *Handbook of Statistical Modelling for the Social and Behavioral Sciences*. New York: Plenum Press.

Clogg, C. C., & Shihadeh, E. S. 1994. *Statistical Models for Ordinal Variables*. Thousand Oaks, CA: Sage.

Clogg, C. C., Rubin, D. B., Schenker, N., Schultz, B., & Weidman, L. 1991. Multiple imputation of industry and occupation codes in census public-use samples using Bayesian logistic-regression. *Journal of the American Statistical Association*, 86, 68-78.

Cochran, W. G. 1950. The comparison of percentages in matched samples. *Biometrika*, 37, 256-266.

Collett, D. 2002. *Modelling Binary Data (Second Edition)*. London: Chapman & Hall/CRC.

Collett, D. 2003. *Modelling Survival Data in Medical Research (Second Edition)*. Boca Raton, FL: Chapman & Hall/CRC.

Congdon, P. 2001. *Bayesian Statistical Modelling*. Chichester: Wiley.

Converse, P. E. 1972. Change in the American electorate. *In*: Campbell, A., & Converse, P. E. (eds), *The Human Meaning of Social Change*. New York, NY: Sage.

Cook, R. D., & Weisberg, S. 1982. *Residuals and Influence in Regression*. London: Chapman & Hall.

Cook, T. D., & Campbell, D. T. 1979. *Quasi-Experimentation*. Boston: Houghton-Mifflin.

Coombs, C. H. 1964. *A Theory of Data*. New York: Wiley.

Corbeil, R. R., & Searle, S. R. 1976. A comparison of variance component estimators. *Biometrics*, 32, 779-791.

Cormack, R. M. 1992. Interval estimation for mark-recapture studies of closed populations. *Biometrics*, 48, 567-576.

Costa, P. T., & McCrae, R. R. 1985. *The NEO Personality Inventory Manual*. Odessa, FL: Psychological Assessment Resources.

Costner, H. L., & Schoenberg, R. 1973. Diagnosing indicator ills in multiple indicator models. *Pages* 167-199 *of*: Goldberger, A. S., & Duncan, O. D. (eds), *Structural Equation Models in the Social Sciences*. New York: Seminar.

Coull, B. A., & Agresti, A. 1999. The use of mixed logit models to reflect heterogeneity in capture-recapture studies. *Biometrics*, 55, 294-301.

Coull, B. A., & Agresti, A. 2000. Random effects modeling of multiple binomial responses using the multivariate binomial logit-normal distribution. *Biometrics*, 56, 73-80.

Cox, D. R. 1961. Tests of separate families of hypotheses. *Pages* 105-123 *of*: Neyman, J. (ed), *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley, CA: University of California Press.

Cox, D. R. 1962. Further results on tests of separate families of hypotheses. *Journal of the Royal Statistical Society, Series B*, 24, 406-424.

Cox, D. R. 1972. Regression models and life tables. *Journal of the Royal Statistical Society, Series B*, 34, 187-203.

Cox, D. R. 1990. Role of models in statistical analysis. *Statistical Science*, 5, 169-174.

Cox, D. R., & Hinkley, D. V. 1974. *Theoretical Statistics*. London: Chapman & Hall.

Cox, D. R., & Oakes, D. 1984. *Analysis of Survival Data*. London: Chapman & Hall.

Cox, D. R., & Solomon, P. J. 2002. *Components of Variance*. Boca Raton, FL: Chapman & Hall/CRC.

Cox, D. R., & Wermuth, N. 1996. *Multivariate Dependencies*. London: Chapman & Hall.

Craig, S. C., & Maggiotto, M. A. 1982. Measuring political efficacy. *Political Methodology*, 8, 85-109.

Crainiceanu, C. M., & Ruppert, D. 2004. Likelihood ratio tests in linear mixed models with one variance component. *Journal of the Royal Statistical Society, Series B*, 66, 165-185.

Crocker, L., & Algina, J. 1986. *Introduction to Classical & Modern Test Theory*. Orlando: Holt, Rinehart and Winston.

Cronbach, L. J. 1971. Test validation. *Pages* 443-507 *of*: Thorndike, R. L. (ed), *Educational Measurement*. Washington: American Council on Education.

Cronbach, L. J., & Meehl, P. E. 1955. Construct validity in psychological tests. *Psychological Bulletin*, 52,281-302.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. 1972. *The Dependability of Behavioral Measurements: Theory of Generalizability for Scores and Profiles*. London: Wiley.

Croon, M. A. 1989. Latent class models for the analysis of rankings. *Pages* 99-121 *of*: De Soete, G., Feger, H., & Klauer, K. C. (eds), *New Developments in Psychological Choice Modeling*. Amsterdam: Elsevier.

Croon, M. A. 2002. Using predicted latent scores in general latent structure models. *Pages* 195-224 *of*: Marconlides, G., & Moustaki, I. (eds), *Latent Variable and Latent Structure Models*. Mahwah, N J: Erlbaum.

Crouch, E. A. C., & Spiegelman, D. 1990. The evaluation of integrals of the form $\int f(t)\exp(-t^2)\,dt$ : Application to logistic-normal models. *Journal of the American Statistical Association*, 85, 464-469.

Crouchley, R., & Davies, R. B. 1999. A comparison of population average and random effects models for the analysis of longitudinal count data with base-line information. *Journal of the Royal Statistical Society*, *Series A*, 162, 331-347.

Crowder, M. J., & Hand, D. J. 1990. *Analysis of Repeated Measures*. London: Chapman & Hall.

Cudeck, R., & Henly, S. 1991. Model selection in covariance structure analysis and the problem of sample size: A clarification. *Psychological Bulletin*, 109, 512-519.

Daganzo, C. 1979, *Multinomial Probit: The Theory and its Application to Demand Forecasting*. New York: Academic.

Dale, J. R. 1986. Global cross-ratio models for bivariate, discrete, ordered responses. *Biometrics*, 42, 909-917.

Danahy, D. T., Burwell, D. T., Aranov, W. S., & Prakash, R. 1976. Sustained hemodynamic and antianginal effect of high dose oral isosorbide dinitrate. *Circulation*, 55, 381-387.

Daniels, M. J., & Gatsonis, C. 1997. Hierarchical polytomous regression models with applications to health services research. *Statistics in Medicine*, 16, 2311-2325.

David, H. A. 1988. *The Method of Paired Comparisons*. Oxford: Oxford University Press.

Davidian, M., & Gallant, A. R. 1992. Smooth nonparametric maximum likelihood estimation for population pharmacokinetics, with application to quindine. *Journal of Pharmacokinetics and Biopharmaceutics*, 20, 529-556.

Davidson, R., & MacKinnon, J. G. 1993. *Estimation and Inference in Econometrics*. Oxford: Oxford University Press.

Davies, R. B. 1993. Statistical modelling for survey analysis. *Journal of the Market Research Society*, 35, 235-247.

Davies, R. B., & Pickles, A. 1987. A joint trip timing store-type choice model for grocery shopping, including inventory effects and nonparametric control for omitted variables. *Transportation Research A*, 21, 345-361.

Davies, R. B., Elias, P., & Penn, R. 1992. The relationship between a husband' s unemployment and his wife' s participation in the labour force. *Oxford Bulletin of Economics and Statistics*, 54, 145-171.

Davis, P. J., & Rabinowitz, P. 1984. *Methods of Numerical Integration*. Orlando: Academic.

Dayton, C. M., & MacReady, G. B. 1988. Concomitant variable latent class models. *Journal of the American Statistical Association*, 83, 173-178.

De Boeck, P., & Wilson, M. 2004. *Explanatory Item Response Models: A Generalized Linear and Nonlinear Approach*. New York: Springer.

de Leeuw, J., & Verhelst, N. D. 1986. Maximum likelihood estimation in generalized Rasch models. *Journal of Educational Statistics*, 11, 183-196.

Deely, J. J., & Lindley, D. V. 1981. Bayes empirical Bayes. *Journal of the American Statistical Association*, 76, 833-841.

DeGroot, M. H. 1970. *Optimal Statistical Decisions*. New York: McGraw-Hill.

Dellaportas, P., & Smith, A. F. M. 1993. Bayesian inference for generalized linear and proportional hazards model via Gibbs sampling. *Journal of the Royal Statistical Society*, *Series C*, 42, 443-460.

Demidenko, E. 2004. *Mixed Models. Theory and Applications*. New York: Wiley.

Dempster, A. P., Laird, N. M., & Rubin, D. B. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, *Series B*, 39, 1-38.

Dempster, A. P., Rubin, D. B., & Tsutakawa, R. D. 1981. Estimation in covariance components models. *Journal of the American Statistical Association*, 76, 341-353.

DerSimonian, R., & Laird, N. M. 1986. Meta-analysis in clinical trials. *Controlled Clinical Trials*, 7, 1777-1788.

Dey, D. K., Gelfand, A. E., Swartz, T. B., & Vlachos, P. K. 1998. Simulation based model checking for hierarchical models. *Test*, 7, 325-346.

Diggle, P. J. 1988. An approach to analysis of repeated measures. *Biometrics*, 44, 959-971.

Diggle, P. J., & Kenward, M. G. 1994. Informative drop-out in longitudinal data analysis (with discussion). *Journal of the Royal Statistical Society*, *Series C*, 43, 49-93.

Digglc, P. J., Heagerty, P. l., Liang, K.-Y., & Zeger, S. L. 2002. *Analysis of Longitudinal Data*. Oxford: Oxford University Press.

Dillon, W. R., Kumar, A., & Deborrero, M. S. 1993. Capturing individual-differences in paired comparisons-an extended BTL model incorporating descriptor variables. *Journal of Marketing Research*, 30, 42-51.

DiPrete, T. A. 2002. Life course risks, mobility regimes and mobility consequences: A comparison of Sweden, Germany, and the United States. *American Journal of Sociology*, 108, 267-309.

Dobbie, M. J., & Welsh, A. H. 2001. Modelling correlated zero-inflated count data. *Australian and New Zealand Journal of Statistics*, 43, 431-444.

Dolan, C. V., & van der Maas, H. L. J. 1998. Fitting multivariate normal finite mixtures subject to structural equation modeling. *Psychometrika*, 63, 227-253.

Du Toit, M. (ed). 2003. *IRT from SSI*. Lincolnwood, IL: Scientific Software International.

DuMouchel, W., Waternaux, C., & Kinney, D. 1996. Hierarchical Bayesian linear models for assessing the effect of extreme

cold weather on schizophrenic births. *In*: Berry, D., & Stangl, D. (eds), *Bayesian Biostatistics*. New York: Marcel Dekker.

Duncan-Jones, P., Grayson, D. A., & Moran, P. A. P. 1986. The utility of latent trait models in psychiatric epidemiology. *Psychological Medicine*, 16, 391-405.

Dunn, G. 1992. Design and analysis of reliability studies. *Statistical Methods in Medical Research*, 1, 123-157.

Dunn, G. 2004. *Statistical Evaluation of Measurement Errors: Design and Analysis of Reliability Studies (Second Edition)*. London: Arnold.

Dunn, G., Everitt, B. S., & Pickles, A. 1993. *Modelling Covariances and Latent Variables using EQS*. London: Chapman & Hall.

Dupacŏvá, J., & Wold, H. 1982. On some identification problems in ML modeling of systems with indirect observation. *Pages* 293-315 *of*: Jöreskog, K. G., & Wold, H. (eds), *Systems under Indirect Observation-Causality ★ Structure ★ Prediction-Part II*. Amsterdam: North-Holland.

Durbin, J. 1951. Incomplete blocks in ranking experiments. *British Journal of Psychology*, 4, 85-90.

Durbin, J., & Watson, G. S. 1950. Testing for serial correlation in least squares regression. *Biometrika*, 37, 409-428.

Ecochard, R., & Clayton, D. G. 2001. Fitting complex random effect models with standard software using data augmentation: Application to a study of male and female fecundability. *Statistical Modelling*, 1, 319-331.

Edwards, J. R., & Bagozzi, R. P. 2000. On the nature and direction of relationships between constructs and measures. *Psychological Methods*, 5, 155-174.

Efron, B. 1977. The efficiency of Cox's likelihood function for censored data. *Journal of the American Statistical Association*, 72, 557-565.

Efron, B., & Hinkley, D. V. 1978. Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information. *Biometrika*, 65, 475-487.

Efron, B., & Morris, C. 1973. Stein's estimation rule and its competitors-an empirical Bayes approach. *Journal of the American Statistical Association*, 68, 117-130.

Efron, B., & Morris, C. 1975. Data analysis using Stein's estimator and its generalizations. *Journal of the American Statistical Association*, 70, 311-319.

Efron, B., &; Tibshirani, R. J. 1993. *An Introduction to the Bootstrap*. London: Chapman & Hall.

EGRET for Windows User Manual. 2000. *EGRET for Windows User Manual*. Cambridge, MA: Citel Software Corporation.

Einhorn, H. 1969. Alchemy in the behavioral sciences. *Public Opinion Quarterly*, 36, 367-378.

Eklöf, M., & Weeks, M. 2003. Estimation of Discrete Choice Models using DCM for Ox: A Manual. *Technical Report*. Downloadable from www. econ. cam. ac. uk/faculty/weeks/DCM/DCMManual/DCMManual, pdf.

Elrod, T., & Keane, M. P. 1995. A factor analytic probit model for representing the market structure in panel data. *Journal of Marketing Research*, 32, 1-16.

Elston, R. C. 1964. On estimating time-response curves. *Biometrics*, 20, 643-647.

Embretson, S. E., & Reise, S. P. 2000. *Item Response Theory for Psychologists*. Mahwah, N J: Erlbaum.

Engel, B., & Keen, A. 1994. A simple approach for the analysis of generalized linear mixed models. *Statistica Neerlandica*, 48, 1-22.

Engle, R. F. 1984. Wald, likelihood-ratio, and Lagrange multiplier tests in econometrics. *Pages* 776-826 *of*: Griliches, Z., & Intriligator, M. D. (eds), *Handbook of Econometrics, Volume II*. Amsterdam: North-Holland.

Engle, R. F., Hendry, D. F., & Richard, J. F. 1983. Exogeneity. *Econometrica*, 51, 277-304.

Etezadi-Amoli, J., & McDonald, R. P. 1983. A second generation nonlinear factor analysis. *Psychometrika*, 48, 315-342.

Evans, M., & Swartz, T. B. 2000. *Approximating Integrals via Monte Carlo and Deterministic Methods*. Oxford: Oxford University Press.

Everitt, B. S. 1987. *Introduction to Optimization Methods and their Application in Statistics*. London: Chapman & Hall.

Everitt, B. S. 1988. A Monte Carlo investigation of the likelihood ratio test for number of classes in latent class analysis. *Multivariate Behsvioral Research*, 23, 531-538.

Everitt, B. S., & Hand, D. J. 1981. *Finite Mixture Distributions*. London: Chapman & Hall.

Everitt, B. S., & Pickles, A. 1999. *Statistical Aspects of the Design and Analysis of Clinical Trials*. London: Imperial College Press.

Ezzet, F., & Whitehead, J. 1991. A random effects model for ordinal responses from a cross-over trial. *Statistics in Medicine*, 10, 901-907.

Fahrmeir, L., & Tutz, G. 2001. *Multivariate Statistical Modelling Based on Generalized Linear Models (Second Edition)*. New York: Springer.

Falconer, D. S. 1981. *Introduction to Quantitative Genetics*. New York: Longman.

Fang, K.-T., & Wang, Y. 1994. *Number-Theoretic Methods in Statistics*. London: Chapman & Hall.

Farewell, V. T. 1979. Some results on the estimation of logistic models based on retrospective data. *Biometrika*, 66, 27-32.

Farewell, V. T., & Prentice, R. L. 1980. The approximation of partial likelihood with emphasis on case-control studies. *Biometrika*, 67, 273-278.

Faucett, C. L., & Thomas, D. C. 1996. Simultaneously modeling censored survival data and repeatedly measured covariates: A Gibbs sampling approach. *Statistics in Medicine*, 15, 1663-1685.

Fayers, P. M., & Hand, D. J. 2002. Causal variables, indicator variables, and measurement scales. *Journal of the Royal*

*Statistical Society, Series A*, 165, 233-261.

Finney, D. J. 1971. *Probit Analysis*. Cambridge: Cambridge University Press.

Fischer, G. H. 1977. Linear logistic trait models: Theory and application. *Pages* 203-225 *of*: Spada, H., & Kempf, W. F. (eds), *Structural Models of Thinking and Learning*. Bern: Huber.

Fischer, G. H. 1995. Derivations of the Rasch model. *Pages* 15-38 *of*: Fischer, G. H., & Molenaar, I. W. (eds), *Rasch models. Foundations, Recent Developments, and Applications*. New York: Springer.

Fisher, F. M. 1966. *The Identification Problem in Econometrics*. New York: McGraw-Hill.

Fisher, R. A. 1922. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London, Series A*, 222, 309-368.

Fitzmaurice, G. M. 1995. A caveat concerning independence estimating equations with multivariate binary data. *Biometrics*, 51,309-317.

Fitzmaurice, G. M., Laird, N. M., & Rotnitzky, A. G. 1993. Regression models for discrete longitudinal responses. *Statistical Science*, 8, 284-309.

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. 2004. *Applied Longitudinal Analysis*. New York: Wiley.

Fleiss, J. L. 1993. The statistical basis of meta-analysis. *Statistical Methods in Medical Research*, 2, 121-145.

Fletcher, R. 1987. *Practical Methods of Optimization*. New York: Wiley.

Fokoué, E., & Titterington, D. M. 2003. Mixtures of factor analysers. Bayesian estimation and inference by stochastic simulation. *Machine Learning*, 50, 73-94.

Follmann, D. A., & Lambert, D. 1989. Generalizing logistic regression by nonparametric mixing. *Journal of the American Statistical Association*, 84, 295-300.

Formann, A. K. 1992. Linear logistic latent class analysis for polytomous data. *Journal of the American Statistical Association*, 87, 476-486.

Formann, A. K., & Kohlmann, T. 1996. Latent class analysis in medical research. *Statistical Methods in Medical Research*, 5, 179-211.

Fornell, C. 1983. Issues in the application of covariance structure analysis: A comment. *Journal of Consumer Research*, 9, 443-450.

Fox, J. P. 2001. *Multilevel IRT: A Bayesian Perspective on Estimating Parameters and Testing Statistical Hypotheses*. Ph. D. thesis, Twente University, Department of measurement and data analysis.

Fox, J. P., & Glas, C. A. W. 2001. Bayesian estimation of a multilevel IRT model using Gibbs sampling. *Psychometrika*, 66, 271-288.

Francis, B. J., Stott, D. N., & Davies, R. B. 1996. *SABRE: A guide for users, version* 3. 1. Lancaster: Centre for Applied Statistics, Lancaster University. Downloadable from http://www. cas. lancs. ac. uk/software/sabre3. 1/sabre. html.

Freedman, D. A. 1983. A note on screening regression equations. *The American Statistician*, 37, 152-155.

Freedman, D. A. 1985. Statistics and the scientific method. *Pages* 345-390 *of*: Mason, M., & Fienberg, S. E. (eds), *Cohort Analysis in Social Research: Beyond the Identification Problem*. New York: Springer.

Freedman, D. A. 1986. As others see us: A case study in path analysis (with discussion). *Journal of Educational Statistics*, 12, 101-128.

Freedman, D. A. 1992. Statistical methods and shoe leather. *Pages* 291-313 *of*: Marsden, P. V. (ed), *Sociological Methodology* 1992. Oxford: Blackwell.

Fruchter, B. 1954. *Introduction to Factor Analysis*. Princeton: Van Nostrand.

Fuller, W. A. 1987. *Measurement Error Models*. New York: Wiley.

Gabriel, K. R. 1962. Ante-dependence analysis of an ordered set of variables. *Annals of Mathematical Statistics*, 33, 201-212.

Gabrielsen, A. 1978. Consistency and identifiability. *Journal of Econometrics*, 8, 261-263.

Gallant, A. R., & Nychka, D. W. 1987. Semi-nonparametric maximum likelihood estimation. *Econometrica*, 55, 363-390.

Gart, J. J., & Zweifel, J. R. 1967. On the bias of various estimators of the logit and its variance, with application to quantal bioassay. *Biometrika*, 54, 181-187.

Geisser, S. 1975. The predictive sample reuse method with applications. *Journal of the American Statistical Association*, 70, 320-328.

Gelman, A., & Rubin, D. B. 1996. Markov chain Monte Carlo methods in biostatistics. *Statistical Methods in Medical Research*, 5, 339-355.

Gelman, A., Goegebeur, Y., Tuerlinckx, F., & Van Mechelen, I. 2000. Diagnostic checks for discrete data regression models using posterior predictive simulations. *Journal of the Royal Statistical Society, Series C*, 49, 247-268.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. 2003. *Bayesian Data Analysis (Second Edition)*. Boca Raton, FL: Chapman & Hall/CRC.

Geman, S., & Geman, D. 1984. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions On Pattern Analysis and Machine Intelligence*, 6, 721-741.

Geraci, V. J. 1976. Identification of simultaneous equation models with measurement error. *Journal of Econometrics*, 4, 262-283.

Geweke, J. 1989. Bayesian inference in econometric models using Monte Carlo integration. *Econometrica*, 57, 1317-1340.

Geweke, J., Keane, M. P., & Runkle, D. 1994. Alternative computational approaches to statistical inference in the

multinomial probit model. *Review of Economics and Statistics*, 76, 609-632.

Geys, H., Molenberghs, G., & Ryan, L. M. 2002. Pseudo-likelihood estimation. *Pages* 90-114 *of*: Aerts, M., Geys, H., Molenberghs, G., & Ryan, L. M. (eds), *Topics in Modelling Clustered Data*. Boca Raton, FL: Chapman & Hall/CRC.

Gibson, W. A. 1959. Three multivariate models: factor analysis, latent structure analysis and latent profile analysis. *Psychometrika*, 24, 229-252.

Gilks, W. R. 1998. Markov chain Monte Carlo. *Pages* 2415-2423 *of*: *Encyclopedia of Biostatistics*. London: Wiley.

Gilks, W. R., & Wild, P. 1992. Adaptive rejection sampling for Gibbs sampling. *Journal of the Royal Statistical Society, Series C*, 41, 337-348.

Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. 1996. Introducing Markov chain Monte Carlo. *Pages* 1-19 *of*: Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (eds), *Markov Chain Monte Carlo in Practice*. Boca Raton: Chapman & Hall/CRC.

Glas, C. A. W. 1988. The derivation of some tests for the Rasch model from the multinomial distribution. *Psychometrika*, 53, 525-546.

Glass, G. V. 1976. Primary, secondary and meta-analysis of research. *Educational Research*, 5, 3-8.

Glass, G. V., & Maguire, T. 1966. Abuses of factor scores. *American Educational Research Journal*, 3, 297-304.

Goldberger, A. S. 1962. Best linear unbiased prediction in the generalized linear regression model. *Journal of the American Statistical Association*, 57, 369-375.

Goldberger, A. S. 1971. Econometrics and psychometrics: A survey of communalities. *Psychometrika*, 36, 83-107.

Goldberger, A. S. 1972. Structural equation methods in the social sciences. *Econometrica*, 40, 979-1001.

Goldberger, A. S. 1991. *A Course in Econometrics*. Cambridge, MA: Harvard University Press.

Goldstein, H. 1986. Multilevel mixed linear model analysis using iterative generalised least squares. *Biometrika*, 73, 43-56.

Goldstein, H. 1987. Multilevel covariance component models. *Biometrika*, 74, 430-431.

Goldstein, H. 1989. Restricted unbiased iterative generalised least squares estimation. *Biometrika*, 76, 622-623.

Goldstein, H. 1991. Nonlinear multilevel models, with an application to discrete response data. *Biometrika*, 78, 45-51.

Goldstein, H. 1994. Recontextualizing mental measurement. *Educational Measurement: Issues and Practice*, 13, 16-19, 43.

Goldstein, H. 2003. *Multilevel Statistical Models (Third Edition)*. London: Arnold.

Goldstein, H., & Browne, W. J. 2002. Multilevel factor analysis modelling using Markov Chain Monte Carlo (MCMC) estimation. *Pages* 225-243 *of*: Marcoulides, G., & Moustaki, I. (eds), *Latent Variable and Latent Structure Models*. Mahwah, N J: Erlbaum.

Goldstein, H., & McDonald, R. P. 1988. A general model for the analysis of multilevel data. *Psychometrika*, 53, 455-467.

Goldstein, H., & Rasbash, J. 1996. Improved approximations for multilevel models with binary responses. *Journal of the Royal Statistical Society, Series A*, 159, 505-513.

Goldstein, H., & Spiegelhalter, D. J. 1996. League tables and their limitations: Statistical issues in comparisons of institutional performance. *Journal of the Royal Statistical Society, Series A*, 159, 385-409.

Goldstein, H., Browne, W. J., & Rasbash, J. 2002. Partitioning variation in multilevel models. *Understanding Statistics*, 1, 223-232.

Goodman, L. A. 1974. Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, 61, 215-231.

Goodman, L. A. 1979. Simple models for the analysis of association in cross-classifications having ordered categories. *Journal of the American Statistical Association*, 74, 537-552.

Goodman, L. A. 1983. The analysis of dependence in cross-classifications having ordered categories, using log-linear models for frequencies and log-linear models for odds. *Biometrics*, 39, 149-160.

Gorsuch, R. L. 1983. *Factor Analysis*. Hillsdale, N J: Erlbaum.

Gould, W., Pitblado, J., & Sribney, W. 2003. *Maximum Likelihood Estimation with Stata*. College Station, TX: Stata Press.

Gourieroux, C., & Monfort, A. 1996. *Simulation-Based Econometric Methods*. Oxford: Oxford University Press.

Gourieroux, C., Monfort, A., Renault, E., & Trognon, A. 1987a. Generalised residuals. *Journal of Econometrics*, 34, 5-32.

Gourieroux, C., Monfort, A., Renault, E., & Trognon, A. 1987b. Simulated residuals. *Journal of Econometrics*, 34, 201-252.

Granger, C. W. J. 2002. Some comments on risk. *Journal of Applied Econometrics*, 17, 447-456.

Green, P. E., & Srinivasan, V. 1990. Conjoint analysis in marketing: New developments with implications for research and practice. *Journal of Marketing*, 54, 3-19.

Greene, V. L., & Carmines, E. G. 1980. Assessing the reliability of linear composites. *Pages* 160-175 *of*: Schuessler, K. F. (ed), *Sociological Methodology* 1980. San Francisco, CA: Jossey-Bass.

Greene, W. H. 2002a. *LIMDEP* 8.0 *Econometric Modeling Guide*. Plainview, NY: Econometric Software.

Greene, W. H. 2002b. *NLOGIT* 3.0 *Reference Guide*. Plainview, NY: Econometric Software.

Greene, W. H. 2003. *Econometric Analysis (Fifth Edition)*. Englewood Cliffs, NJ: Prentice Hall.

Greenland, S., & Brumback, B. 2002. An overview of relations among causal modelling methods. *International Journal of Epidemiology*, 31, 1030-1037.

Greenwood, M., & Yule, G. U. 1920. An inquiry into the nature of frequency distributions of multiple happenings, with particular reference to the occurence of multiple attacks of disease or repeated accidents. *Journal of the Royal Statistical*

*Society, Series A*, 83, 255-279.

Griliches, Z. 1974. Errors in variables and other unobservables. *Econometrica*, 42, 971-998.

Grilli, L., & Rampichini, C. 2003. Alternative specifications of bivariate multilevel probit ordinal response models. *Journal of Educational and Behavioral Statistics*, 28, 31-44.

Gueorguieva, R., & Sanacora, G. 2003. A latent variable model for joint analysis of repeatedly measured ordinal and continuous outcomes. *Pages* 171-176 *of*: Verbeke, G., Molenberghs, C., Aerts, A., & Fieuws, S. (eds), *Proceedings of the 18th International Workshop on Statistical Modelling.* Leuven, Belgium: Katholieke Universiteit Leuven.

Gulliksen, H. 1950. *Theory of Mental Tests.* New York: Wiley.

Gustafson, P. 2004. *Measurement Error and Misclassification in Statistics and Epidemiology. Impacts and Bayesian Adjustments.* Boca Raton, FL: Chapman & Hall/CRC.

Guttmann, L. 1955. The determinacy of factor score matrices with implications for five other basic problems of common factor theory. *British Journal of Mathematical and Statistical Psychology*, 8, 65-81.

Guttmann, L. 1977. What is not what in statistics. *Journal of the Royal Statistical Society, Series D*, 26, 81-107.

Haaijer, M. E., Wedel, M., Vriens, M., & Wansbeek, T. J. 1998. Utility covariances and context effects in conjoint MNP models. *Marketing Science*, 17, 236-252.

Haberman, S. J. 1977. Maximum likelihood estimates in exponential response models. *The Annals of Statistics*, 5, 815-841.

Haberman, S. J. 1979. *Analysis of Qualitative Data: New Developments.* Vol. 2. New York: Academic.

Haberman, S. J. 1989. A stabilized Newton-Raphson algorithm for log-linear models for frequency tables derived by indirect observation. *Pages* 193-211 *of*: Clogg, C. C. (ed), *Sociological Methodology* 1988. Washington: American Sociological Association.

Hagenaars, J. A. 1988. Latent structure models with direct effects between indicators: Local dependence models. *Sociological Methods & Research*, 16, 379-405.

Hagenaars, J. A. 1993. *Loglinear Models with Latent Variables.* Sage University Paper Series on Quantitative Applications in the Social Sciences. Newbury Park, CA: Sage. Hajivassiliou, V. A., & Ruud, P. A. 1994. Classical estimation methods for LDV models using simulation. *Pages* 2383-2441 *of*: Engle, R. F., & McFadden, D. L. (eds), *Handbook of Econometrics*, vol. IV. New York: Elsevier.

Hajivassiliou, V. A., McFadden, D. L., & Ruud, P. A. 1996. Simulation of multivariate normal rectangle probabilities and their derivatives - Theoretical and computational results. *Journal of Econometrics*, 72, 85-134.

Hall, D. B. 1997. Zero-inflated Poisson and binomial regression with random effects: A case study. *Biometrics*, 56, 1030-1039.

Halperin, S. 1976. The incorrect measurement of components. *Educational and Psychological Measurement*, 36, 347-353.

Hambleton, R. K., & Swaminathan, H. 1985. *Item Response Theory: Principles and Applications.* Boston: Kluwer.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. 1991. *Fundamentals of Item Response Theory.* Newbury Park, CA: Sage.

Hamerle, A., & Rönning, G. 1995. Panel analysis for qualitative variables. *Pages* 401-451 *of*: Arminger, G., Clogg, C. C., & Sobel, M. E. (eds), *Handbook of Statistical Modelling for the Social and Behavioral Sciences.* New York: Plenum Press.

Han, A., & Hausman, J. A. 1990. Flexible parametric estimation of duration and competing risk models. *Journal of Applied Econometrics*, 5, 1-29.

Hand, D. J., & Crowder, M. J. 1996. *Practical Longitudinal Data Analysis.* London: Chapman & Hall.

Hardin, J., & Hilbe, J. 2002. *Generalized Estimating Equations.* Boca Raton, FL: Chapman & Hall/CRC.

Harman, H. H. 1976. *Modern Factor Analysis.* Chicago: University of Chicago Press.

Harper, D. 1972. Local dependence latent structure models. *Psychometrika*, 37, 53-59.

Harrell, F. E. 2001. *Regression Modeling Strategies. With Application to Linear Models, Logistic Regression and Survival Analysis.* New York: Springer.

Harris, C. W. 1967. On factors and factor scores. *Psychometrika*, 32, 363-379.

Hartzel, J., Agresti, A., & Caffo, B. 2001. Multinomial logit random effects models. *Statistical Modelling*, 1, 81-102.

Harvey, A. 1976. Estimating regression models with multiplicative heteroscedasticity. *Econometrica*, 44, 461-465.

Harville, D. A. 1977. Maximum likelihood approaches to variance components estimation and related problems. *Journal of the American Statistical Association*, 72, 320-340.

Hauck, W. W., & Donner, A. 1977. Wald's test as applied to hypotheses in logit analysis. *Journal of the American Statistical Association*, 72, 851-853.

Hauser, R. M., & Goldberger, A. S. 1971. The treatment of unobservable variables in path analysis. *Pages* 81-117 *of*: Costner, H. L. (ed), *Sociological Methodology* 1971. San Francisco, CA: Jossey-Bass.

Hausman, J. A. 1978. Specification tests in econometrics. *Econometrica*, 46, 1251-1271.

Hausman, J. A., & Ruud, P. A. 1987. Specifying and testing econometric models for rankordered data. *Journal of Econometrics*, 34, 83-103.

Hausman, J. A., & Wise, D. A. 1978. A conditional probit model for qualitative choice: Discrete decisions recognizing interdependence and heterogeneous preferences. *Econometrica*, 46, 403-426.

Hausman, J. A., & Wise, D. A. 1979. Attrition bias in experimental and panel data: The Gary income maintenance experiment. *Econometrica*, 47, 455-473.

Hausman, J. A., Hall, B. H., &; Griliches, Z. 1984. Econometric models for count data with an application to the patents R and D relationship. *Econometrica*, 52, 909-938.

Heagerty, P. J., & Zeger, S. L. 2000. Marginalized multilevel models and likelihood inference. *Statistical Science*, 15, 1-26.

Heath, A., Curtice, J. K., Jowell, R., Evans, G., Fields, J., & Witherspoon, S. 1991. *Understanding Political Change: The British Voter* 1964-1987. Oxford: Pergamon Press.

Heath, A., Jowell, R., Curtice, J. K., Brand, J. A., & Mitchell, J. C. 1993. *British Election Panel Study*, 1987-1992 [*Computer file*] *SN*: 2983. Colchester, Essex: The Data Archive [Distributor].

Heath, A., Jowell, R., & Curtice, J. K. (eds). 1994. *Labour's Last Chance? The* 1992 *Election and Beyond.* Aldershot: Darthmouth Publishing Company.

Heckman, J. J. 1978. Dummy endogenous variables in a simultaneous equation system. *Econometrica*, 46, 931-959.

Heckman, J. J. 1979. Sample selection bias as a specification error. *Econometrica*, 47, 153-161.

Heckman, J. J. 1981a. Heterogeneity and state dependence. *Pages* 91-139 *of*: Rosen, S. (ed), *Studies in Labor Markets.* Chicago: Chicago University Press.

Heckman, J. J. 1981b. The incidental parameters problem and the problem of initial conditions in estimating a discrete stochastic process and some Monte Carlo evidence on their practical importance. *Pages* 179-196 *of*: Manski, C. F., & McFadden, D. L. (eds), *Structural Analysis of Discrete Data with Econometric Applications.* Cambridge: MIT Press.

Heckman, J. J. 1981c. Statistical models for discrete panel data. *Pages* 114-178 *of*: Manski, C. F., & McFadden, D. L. (eds), *Structural Analysis of Discrete Data with Econometric Applications.* Cambridge: MIT Press.

Heckman, J. J., & Singer, B. 1982. The identification problem in econometric models for duration data. *Pages* 39-77 *of*: Hildenbrand, W. (ed), *Advances in Econometrics*, *Proceedings of the Fourth World Congress of the Econometric Society.* Cambridge: Cambridge University Press.

Heckman, J. J., & Singer, B. 1984. A method of minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica*, 52, 271-320.

Heckman, J. J., & Willis, R. J. 1977. A beta-logistic model for the analysis of sequential labor force participation by married women. *Journal of Political Economy*, 85, 27-58.

Hedeker, D. 1999. MIXNO: A computer program for mixed-effects logistic regression. *Journal of Statistical Software*, 1-92.

Hedeker, D. 2003. A mixed-effects multinomial logistic regression model. *Statistics in Medicine*, 22, 1433-1466.

Hedeker, D., & Gibbons, R. D. 1994. A random-effects ordinal regression model for multilevel analysis. *Biometrics*, 50, 933-944.

Hedeker, D., & Gibbons, R. D. 1996a. MIXOR: A computer program for mixed-effects ordinal probit and logistic regression analysis. *Computer Methods and Programs in Biomedicine*, 49, 157-176.

Hedeker, D., & Gibbons, R. D. 1996b. MIXREG: A computer program for mixed-effects regression analysis with autocorrelated errors. *Computer Methods and Programs in Biomedicine*, 49, 229-252.

Hedeker, D., Siddiqui, O., & Hu, F. B. 2000. Random-effects regression analysis of correlated grouped-time survival data. *Statistical Methods in Medical Research*, 9, 161-179.

Hedges, L. V., & Olkin, I. 1985. *Statistical Methods for Meta-analysis.* Orlando, FL: Academic Press.

Heinen, T. 1996. *Latent Class and Discrete Latent Trait Models: Similarities and Differences.* Thousand Oaks, CA: Sage.

Henderson, C. R. 1975. Best linear unbiased estimation and prediction under a selection model. *Biometrics*, 31,423-447.

Hershberger, S. L. 1994. The specification of equivalent models before the collection of data. *Pages* 68-108 *of*: von Eye, A., & Clogg, C. C. (eds), *Latent Variable Analysis: Applications to Developmental Research.* Thousand Oaks, CA: Sage.

Hilden-Minton, J. A. 1995. *Multilevel Diagnostics for Mixed and Hierarchical Linear Models.* Ph. D. thesis, Department of Mathematics, University of California, Los Angeles.

Hildesheim, A., Mann, V., Brinton, L. A., Szklo, M., Reeves, W. C., & Rawls, W. E. 1991. Herpes simplex virus type 2: A possible interaction with human papillomavirus types 16/18 in the development of invasive cervical cancer. *International Journal of Cancer*, 49, 335-340.

Hill, P. W., & Goldstein, H. 1998. Multilevel modeling of educational data with cross-classification and missing identification of units. *Journal of Educational and Behavioral Statistics*, 23, 117-128.

Hobcraft, J., & Murphy, M. 1987. Demographic event history analysis: a selective review. *In*: Crouchley, R. (ed), *Longitudinal Data Analysis.* Aldershot: Avebury.

Hobert, J. P. 2000. Hierarchical models: A current computational perspective. *Journal of the American Statistical Association*, 95, 1312-1316.

Hodges, J. S., & Sargent, D. J. 2001. Counting degrees of freedom in hierarchical and other richly-parameterised models. *Biometrika*, 88, 367-379.

Hogan, J. W., & Laird, N. M. 1997a. Mixture models for the joint distribution of repeated measures and event times. *Statistics in Medicine*, 16, 239-257.

Hogan, J. W., & Laird, N. M. 1997b. Model-based approaches to analyzing incomplete repeated measures and failure time data. *Statistics in Medicine*, 16, 259-271.

Hoijtink, H. 2001. Confirmatory latent class analysis: model selection using Bayes factors and (pseudo) likelihood ratio statistics. *Multivariate Behavioral Research*, 36, 563-588.

Hoijtink, H., & Boomsma, A. 1995. On person parameter estimation in the dichotomous Rasch model. *Pages* 53-68 *of*: Fischer, G. H., & Molenaar, I. W. (eds), *Rasch Models: Foundations, Recent Developments, and Applications.* New York: Springer.

Holford, T. R. 1976. Life tables with concomitant variables. *Biometrics*, 32, 587-597.

Holford, T. R. 1980. The analysis of rates and survivorship using log-linear models. *Biometrics*, 36, 299-305.

Holford, T. R., White, C., & Kelsey, J. L. 1978. Multivariate analysis for matched casecontrol studies. *American Journal Epidemiology*, 107, 245-256.

Holland, P. W. 1988. Causal inference, path analysis, and recursive structural equation models. *Pages* 449-493 *of*: Clogg, C. C. (ed), *Sociological Methodology* 1988. Washington: American Sociological Association.

Holmås, T. H. 2002. Keeping nurses at work: A duration analysis. *Health Economics*, 11, 493-503.

Holzinger, K. L., & Harman, H. H. 1941. *Factor Analysis: A Synthesis of Factorial Methods.* Chicago: University of Chicago Press.

Horowitz, J. L., Sparmann, J. M., & Daganzo, C. 1982. An investigation of the accuracy of the Clark approximation for the multinomial probit model. *Transportation Science*, 16, 382-401.

Hosmer, D. A., & Lemeshow, S. A. 1999. *Applied Survival Analysis: Regression Modeling of Time to Event Data.* New York: Wiley.

Hosmer, D. A., & Lemeshow, S. A. 2000. *Applied Logistic Regression (Second Edition).* New York: Wiley.

Hotelling, H. 1957. The relations of the newer multivariate statistical methods to factor analysis. *British Journal of Statistical Psychology*, 10, 69-79.

Hougaard, P. 1986a. A class of multivariate failure time distributions. *Biometrika*, 73, 671-678.

Hougaard, P. 1986b. Survival models for heterogeneous populations derived from stable distributions. *Biometrika*, 73, 387-396.

Hougaard, P. 2000. *Analysis of Multivariate Survival Data.* New York: Springer.

Hox, J. 2002. *Multilevel Analysis: Techniques and Applications.* Mahwah, NJ: Erlbaum.

Hsiao, C. 1983. Identification. *Pages* 223-283 *of*: Griliches, Z., & Intriligator, M. D. (eds), *Handbook of Econometrics, Volume I.* Amsterdam: North-Holland.

Hsiao, C. 2002. *Analysis of Panel Data (Second Edition).* Cambridge, UK: Cambridge University Press.

Hu, P., Tsiatis, A. A., & Davidian, M. 1998. Estimating the parameters in the Cox model when the covariate variables are measured with error. *Biometrics*, 54, 1407-1419.

Huber, P. J. 1967. The behavior of maximum likelihood estimates under non-standard conditions. *Pages* 221-233 *of*: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, vol. I. Berkeley: University of California Press.

Huggins, R. M. 1989. On the statistical analysis of capture experiments. *Biometrika*, 76, 133-140.

Hunt, M. 1997. *How Science Takes Stock: The Story of Meta-analysis.* New York, NY: Sage.

Imbens, G. W., & Rubin, D. B. 1997a. Bayesian inference for causal effects in randomized experiments with noncompliance. *The Annals of Statistics*, 25, 305-327.

Imbens, G. W., & Rubin, D. B. 1997b. Estimating outcome distributions for compliers in instrumental variable models. *Review of Economic Studies*, 64, 555-574.

Inglehart, R. 1977. *The Silent Revolution: Changing Values and Political Styles Among Western Publics.* Princeton, NJ: Princeton University Press.

James, L. R., Mulaik, S. A., & Brett, J. M. 1982. *Causal Analysis: Assumptions, Models, and Data.* Beverly Hills, CA: Sage.

James, W., & Stein, C. 1961. Estimation with quadratic loss. *Pages* 361-379 *of*: *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, vol. I. Berkeley: University of California Press.

Jedidi, K., Jagpal, H. S., & DeSarbo, W. S. 1997. Finite-mixture structural equation models for response-based segmentation and unobserved heterogeneity. *Marketing Science*, 16, 39-59.

Jenkins, S. P. 1995. Easy estimation methods for discrete-time duration models. *Oxford Bulletin of Economics and Statistics*, 57, 129-137.

Jewell, N. P. 1982. Mixtures of exponential distributions. *The Annals of Statistics*, 10, 479-484.

Jo, B. 2002. Model misspecification sensitivity in estimating causal effects of interventions with noncompliance. *Statistics in Medicine*, 21, 3161-3181.

Johansen, S. 1983. An extension of Cox's regression model. *International Statistical Review*, 51, 258-262.

Johnson, N. L., Kotz, S., &; Balakrishnan, N. 1994. *Distributions in Statistics: Continuous Univariate Distributions.* Vol. 1. New York: Wiley.

Johnson, R. A., & Wichern, D. W. 1983. *Applied Multivariate Statistical Analysis.* Englewood Cliffs: Prentice-Hall.

Johnson, V. E., & Albert, J. H. 1999. *Ordinal Data Modelling.* New York: Springer.

Joliffe, I. 1986. *Principal Components Analysis.* New York: Springer.

Jones, R. H. 1993. *Longitudinal Data with Serial Correlation: A State-Spece Approach.* London: Chapman and Hall.

Jöreskog, K. G. 1967. Some contributions to maximum likelihood factor analysis. *Psychometrika*, 32, 443-482.

Jöreskog, K. G. 1969. A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34, 183-202.

Jöreskog, K. G. 1971a. Simultaneous factor analysis in several populations. *Psychometrika*, 36, 409-426.

Jöreskog, K. G. 1971b. Statistical analysis of sets of congeneric tests. *Psychometrika*, 36, 109-133.

Jöreskog, K. G. 1973. A general method for estimating a linear structural equation system. *Pages* 85-112 *of*: Goldberger, A. S., & Duncan, O. D. (eds), *Structural Equation Models in the Social Sciences*. New York: Seminar.

Jöreskog, K. G. 1974. Analysing psychological data by structural analysis of covariance matrices. *Pages* 1-56 *of*: Krantz, D. H., Atkinson, R. C., Luce, R. D., & Suppes, P. (eds), *Contemporary Developments in Mathematical Psychology*, vol. II. San Francisco, CA: Freeman.

Jöreskog, K. G. 1978. Structural analysis of covariance and correlation matrices. *Psychometrika*, 43, 443-477.

Jöreskog, K. G. 1993. Testing structural equation models. *Pages* 294-316 *of*: Bollen, K. A., & Long, J. S. (eds), *Testing Structural Equation Models*. Newbury Park, CA: Sage.

Jöreskog, K. G. 1998. Interaction and nonlinear modeling. *Pages* 239-250 *of*: Schumaker, R. E., & Marcoulides, G. A. (eds), *Interaction and Nonlinear Effects in Structural Equation Modeling*. Mahwah, N J: Erlbaum.

Jöreskog, K. G., & Sörbom, D. 1989. *LISREL 7: A Guide to the Program and Applications*. Chicago, IL: SPSS Publications.

Jöreskog, K. G., & Sörbom, D. 1990. Model search with TETRAD and LISREL. *Sociological Methods & Research*, 19, 93-106.

Jöreskog, K. G., & Sörbom, D. 1994. *LISREL 8 and PRELIS 2: Comprehensive Analysis of Linear Relationships in Multivariate Data. LISREL 8 User's Guide*. Hillsdale, NJ: Erlbaum.

Jöreskog, K. G., Sörbom, D., Du Toit, S. H. C., & Du Toit, M. 2001. *LISREL 8: New Statistical Features*. Lincolnwood, IL: Scientific International.

Judge, G. G., Griffiths, W. E., Hill, R. C., Lütkepol, H., & Lee, T.-C. 1985. *The Theory and Practice of Econometrics*. New York: Wiley.

Kalbfleisch, J. D., & Prentice, R. L. 2002. *The Statistical Analysis of Failure Time Data (Second Edition)*. New York: Wiley.

Kamakura, W. A., Kim, B. D., & Lee, J. 1996. Modeling preference and structural heterogeneity in consumer choice. *Marketing Science*, 15, 152-172.

Kass, R. E., & Raftery, A. E. 1995. Bayes factors. *Journal of the American Statistical Association*, 90, 773-795.

Kass, R. E., & Steffey, D. 1989. Approximate Bayesian inference in conditionally independent hierarchical models (parametric empirical Bayes models). *Journal of the American Statistical Association*, 84, 717-726.

Kass, R. E., & Wasserman, L. 1995. A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *Journal of the American Statistical Association*, 90, 928-934.

Katz, E. 2001. Bias in conditional and unconditional fixed effects logit estimation. *Political Analysis*, 9, 379-384.

Keane, M. P. 1992. A note on identification in the multinomial probit model. *Journal of Business and Economic Statistics*, 10, 193-200.

Keane, M. P. 1994. A computationally practical simulation estimator for panel data. *Econometrica*, 62, 95-116.

Keiding, N. 1992. Independent delayed entry (with discussion). *Pages* 309-326 *of*: Klein, J. P., & Goel, P. K. (eds), *Survival Analysis: State of the Art*. Dordrecht: Kluwer.

Kelly, P. J., & Lim, L. L-Y. 2000. Survival analysis for recurrent event data: An application to childhood infectious diseases. *Statistics in Medicine*, 19, 13-33.

Kemp, I., Boyle, P., Smans, M., & Muir, C. 1985. *Atlas of Cancer in Scotland*, 1975-1980. *Incidence and Epidemiologic Perspective*. Scientific Publication 72. Lyon, France: International Agency for Research on Cancer.

Kempthorne, O. 1980. The term "design matrix" *The American Statistician*, 34, 249.

Kendall, M., & Stuart, A. 1979. *The Advanced Theory of Statistics* . Vol. II. New York: Macmillan.

Kenkel, D. S., & Terza, J. V. 2001. The effect of physician advice on alcohol consumption: count regression with an endogenous treatment effect. *Journal of Applied Econometrics*, 16, 165-184.

Kenny, D. A., & Judd, C. M. 1984. Estimating the non-linear and interactive effects of latent variables. *Psychological Bulletin*, 96, 201-210.

Kestelman, H. 1952. The fundamental equation of factor analysis. *British Journal of Psychology*, *Statistical Section*, 5, 1-6.

Kim, J.-O., & Mueller, C. W. 1978. *Factor Analysis. Statistical Methods and Practical Issues*. Sage University Paper Series on Quantitative Applications in the Social Sciences. Beverly Hills, CA: Sage.

Kim, J.-O., & Rabjohn, J. 1978. Binary items and index construction. *Pages* 120-159 *of*: Schussler, K. F. (ed), *Sociological Methodology* 1978. San Francisco, CA: Jossey-Bass.

King, G., Murray, C. J. L., Salomon, J. A., & Tandon, A. 2003. Enhancing the validity of cross-cultural comparability of survey research. *American Political Science Review*, 97, 567-583.

Kirisci, L., & Hsu, T.-C. 2001 Robustness of item parameter estimation programs to assumptions of unidimensionality and normality. *Applied Psychological Measurement*, 25. 146-162.

Klecka, W. R. 1980. *Discriminant Analysis*. Sage University Paper Series on Quantitative Applications in the Social Sciences. Beverly Hills, CA: Sage.

Klein, J. P., & Moeschberger, M. L. 2003. *Survival Analysis: Techniques for Censored and Truncated Data (Second Edition)*. New York: Springer.

Knorr-Held, L., & Best, N. G. 2001. A shared component model for detecying joint and selective clustering of two diseases. *Journal of the Royal Statistical Society, Series A*, 164, 73-85.

Knott, M., Albanese, M. T., & Galbraith, J. I. 1990. Scoring attitudes to abortion. *Journal of the Royal Statistical Society, Series D*, 40, 217-223.

Koopmans, T. C., & Reiersø l, O. 1950. The identification of structural characteristics. *Annals of Mathematical Statistics*, 21, 165-181.

Krane, W. R., & McDonald, R. P. 1978. Scale invariance and the factor analysis of correlation matrices. *British Journal of Mathematical and Statistical Psychology*, 31, 218-228.

Kreft, I. G. C., & de Leeuw, J. 1994. The gender gap in earnings. *Sociological Methods & Research*, 22, 319-341.

Kreft, I. G. G., & de Leeuw, J. 1998. *Introducing Multilevel Modeling.* London: Sage.

Kruskal, J. B., & Wish, M. 1978. *Multidimensional Scaling.* Sage University Paper Series on Quantitative Applications in the Social Sciences. Beverly Hills, CA: Sage.

Kuk, A. Y. C. 1995. Asymptotically unbiased estimation in generalized linear models with random effects. *Journal of the Royal Statistical Society, Series B*, 57, 395-407.

Kuk, A. Y. C. 1999. Laplace importance sampling for generalized linear mixed models. *The Journal of Statistical Computation and Simulation*, 63, 143-158.

Kuk, A. Y. C., & Cheng, Y. W. 1997. The Monte Carlo Newton-Raphson algorithm. *The Journal of Statistical Computation and Simulation*, 59, 233-250.

Küsters, U. 1987. *Hierarchische Mittelwert- und Kovarianzstrukturmodelle mit Nichtmetrischen Endogenen Variablen.* Heidelberg: Physica.

Kvalem, I. L., & Træen, B. 2000. Self-efficacy, scripts of love and intention to use condoms among Norwegian adolescents. *Journal of Youth and Adolescents*, 29, 337-353.

Lader, D., & Matheson, J. 1991. *Smoking among Secondary School Children in* 1990. London: HMSO.

Laird, N. M. 1978. Nonparametric maximum likelihood estimation of a mixing distribution. *Journal of the American Statistical Association*, 73, 805-811.

Laird, N. M. 1982. Empirical Bayes estimates using the nonparametric maximum likelihood estimate for the prior. *The Journal of Statistical Computation and Simulation*, 15, 211-220.

Laird, N. M. 1988. Missing data in longitudinal studies. *Statistics in Medicine*, 7, 305-315.

Laird, N. M., & Olivier, D. 1981. Covariance analysis of censored survival data using log-linear analysis techniques. *Journal of the American Statistical Association*, 76, 231-240.

Laird, N. M., & Ware, J. H. 1982. Random effects models for longitudinal data. *Biometrics*, 38, 963-974.

Lambert, D. 1992. Zero-inflated Poisson-regression with an application to defects in manufacturing. *Technometrics*, 34, 1-14.

Lancaster, T. 1990. *The Econometric Analysis of Transition Data.* Cambridge: Cambridge University Press.

Lancaster, T. 2000. The incidental parameter problem since 1948. *Journal of Econometrics*, 95, 391-413.

Lange, N., & Ryan, L. M. 1989. Assessing normality in random effects models. *The Annals of Statistics*, 17, 624-642.

Langford, I. H., & Lewis, T. 1998. Outliers in multilevel data. *Journal of the Royal Statistical Society, Series A*, 161, 121-160.

Langford, I. H., Leyland, A. H., Rasbash, J., & Goldstein, H. 1999. Multilevel modeling of the geographical distributions of rare diseases. *Journal of the Royal Statistical Society, Series C*, 48, 253-268.

Lawley, D. N. 1942. Further investigations in factor estimation. *Pages* 176-185 *of*: *Proceedings of the Royal Society of Edinburgh*, vol. 61.

Lawley, D. N., & Maxwell, A. E. 1971. *Factor Analysis as a Statistical Method.* London: Butterworths.

Lawson, A. B., Browne, W. J., & Vidal-Rodeiro, C. L. 2003. *Disease Mapping with WinBUGS and MLwiN.* New York: Wiley.

Lazarsfeld, P. F. 1950. The logical and mathematical foundation of latent structure analysis. *Pages* 362-412 *of*: Stouffer, S. A., Guttmann, L., Suchman, E. A., Lazarsfeld, P. F., Star, S. A., & Clausen, J. A. (eds), *Studies in Social Psychology in World War II*, vol. 4, Measurement and Prediction. Princeton, N J: Princeton University Press.

Lazarsfeld, P. F. 1959. Latent structure analysis. *In*: Koch, S. (ed), *Psychology: A Study of A Science, vol.* III. New York: McGraw-Hill.

Le Cessie, S., & van Houwelingen, J. C. 1994. Logistic regression for correlated binary data. *Journal of the Royal Statistical Society, Series C*, 43, 95-108.

Leamer, E. E. 1978. *Specification Searches.* New York: Wiley.

Lee, E. W., Wei, L. J., & Amato, D. A. 1992. Cox-type regression analysis for large numbers of small groups of correlated failure time observations. *Pages* 237-247 *of*: Klein, J. P., & Goel, P. K. (eds), *Survival Analysis: State of the Art.* Dordrecht: Kluwer.

Lee, L.-F. 1982. Health and wage: A simultanous equation model with multiple discrete indicators. *International Economic Review*, 23, 199-221.

Lee, S.-Y., & Shi, J.-Q. 2001. Maximum likelihood estimation of two-level latent variable models with mixed continuous and polytomous data. *Biometrics*, 57, 787-794.

Lee, Y., & Nelder, J. A. 1996. Hierarchical generalized linear models (with discussion). *Journal of the Royal Statistical Society, Series B*, 58, 619-678.

Lee, Y., & Nelder, J. A. 2001. Hierarchical generalised linear models: A synthesis of generalised linear models, random-effect models and structured dispersions. *Biometrika*, 88, 987-1006.

Lehmann, E. L. 1990. Model specification: The views of Fisher and Neyman, and later developments. *Statistical Science*, 5, 160-168.

Lenk, P. J., & DeSarbo, W. S. 2000. Bayesian inference for finite mixtures of generalized linear models with random effects. *Psychometrika*, 65, 93-119.

Leppik, I. E., Dreifuss, F. E., Porter, R., Bowman, T., Santilli, N., Jacobs, M., Crosby, C., Cloyd, J., Stackman, J., Graves, N., Sutula, T., Welty, T., Vickery, J., Brundage, R., Gates, J., Gumnit, R. J., & Guttierrez, A. 1987. A controlled-study of progabide in partial seizures - methodology and results. *Neurology*, 37, 963-968.

Lerman, S., & Manski, C. F. 1981. On the use of simulated frequencies to approximate choice probabilities. *In*: Manski, C. F., & McFadden, D. L. (eds), *Structural Analysis of Discrete Data with Econometric Applications.* Cambridge, MA: MIT Press.

Lesaffre, E., & Spiessens, B. 2001. On the effect of the number of quadrature points in a logistic random-effects model: An example. *Journal of the Royal Statistical Society*, *Series C*, 50, 325-335.

Lesaffre, E., & Verbeke, G. 1998. Local influence in linear mixed models. *Biometrics*, 54, 570-582.

Lewis, T., & Langford, I. H. 2001. Outliers, robustness and detection of discrepant data. *Pages* 75-90 *of*: Leyland, A. H., & Goldstein, H. (eds), *Multilevel Modelling of Health Statistics.* Chichester, UK: Wiley.

Leyland, A. H. 2001. Spatial analysis. *Pages* 127-140 *of*: Leyland, A. H., & Goldstein, H. (eds), *Multilevel Modelling of Health Statistics.* Chichester: Wiley.

Liang, K.-Y., & Zeger, S. L. 1986. Longitudinal data analysis using generalized linear models. *Biometrika*, 73, 13-22.

Liang, K.-Y., Zeger, S. L., & Qaqish, B. 1992. Multivariate regression analysis for categorical data. *Journal of the Royal Statistical Society*, *Series B*, 54, 3-40.

Likert, R. 1932. A technique for the measurement of attitudes. *Archives of Psychology*, 140. 5-53.

Lillard. L. A 1993. Simultaneous-equations for hazards-marriage duration and fertility timing *Journal of Econometrics*, 56, 189-217.

Lillard, L. A., & Panis, C. W. A. 2000. *aML User's Guide and Reference Manual.* Los Angeles, CA: EconWare.

Lin, X., & Breslow, N. E. 1996. Bias correction in generalized linear mixed models with multiple components of dispersion. *Journal of the American Statistical Association*, 91. 1007-1016.

Linda, N. Y., Lee, S.-Y., & Poon, W.-Y. 1993. Covariance structure analysis with three level data. *Computational Statistics & Data Analysis*, 15, 159-178.

Lindley, D. V. 1969. Discussion of Copas (1969): Compound decisions and empirical Bayes. *Journal of the Royal Statistical Society*, *Series B*, 31, 419-421.

Lindley, D. V. 1971. The estimation of many parameters. *Pages* 435-455 *of*: Godambe, V. P., & Sprott, D. A. (eds), *Foundations of Statistical Inference.* Toronto: Holt, Rinehart and Winston.

Lindley, D. V., & Smith, A. F. M. 1972. Bayes estimates for the linear model (with discussion). *Journal of the Royal Statistical Society*, *Series B*, 34, 1-41.

Lindsay, B. G. 1983. The Geometry of Mixture Likelihoods: A General Theory. *The Annals of Statistics*, 11, 86-94.

Lindsay, B. G. 1995. *Mixture Models*: *Theory*, *Geometry and Applications.* NSF-CBMS Regional Conference Series in Probability and Statistics, vol. 5. Hayward, CA: Institute of Mathematical Statistics.

Lindsay, B. G., Clogg, C. C., & Grego, J. 1991. Semiparametric estimation in the Rasch model and related exponential response models, including a simple latent class model for item analysis. *Journal of the American Statistical Association*, 86, 96-107.

Lindsey, J. K. 1996. *Parametric Statistical Inference.* Oxford, UK: Oxford University Press.

Lindsey, J. K. 1999. *Models for Repeated Measurements* (*Second Edition*). Oxford, UK: Oxford University Press.

Lindsey, J. K., & Lambert, P. 1998. On the appropriateness of marginal models for repeated measurements in clinical trials. *Statistics in Medicine*, 17, 447-469.

Lipsitz, S., Laird, N. M., & Harrington, D. 1991. Generalized estimating equations for correlated binary data: Using odds ratios as a measure of association. *Biometrika*, 78, 153-160.

Listhaug, O. 1989. *Citizens, Parties and Norwegian Electoral Politics* 1957-1985. Trondheim: Tapir.

Little, R. J. A. 1995. Modeling the drop-out mechanism in repeated measures studies. *Journal of the American Statistical Association*, 90, 1112-1121.

Little, R. J. A., & Rubin, D. B. 1983. Models for nonresponse in sample surveys. *The American Statistician*, 37, 218-220.

Little, R. J. A., N Rubin, D. B. 2002. *Statistical Analysis with Missing Data* (*Second Edition*). New York: Wiley.

Little, R. J. A., & Yau, L. H. Y. 1998. Statistical techniques for analyzing data from prevention trials. *Psychological Methods*, 3, 147-159.

Liu, C., Rubin, D. B., & Wu, Y. 1998. Parameter expansion to accelerate EM: The PX-EM algorithm. *Biometrika*, 85, 755-770.

Liu, Q., & Pierce, D. A. 1994. A note on Gauss-Hermite quadrature. *Biometrika*, 81, 624-629.

Lohmöller, J. 1989. *Latent Variable Path Modeling with Partial Least Squares.* Heidelberg: Physica-Verlag.

Long, J. S. 1997. *Regression Models for Categorical and Limited Dependent Variables.* Thousand Oaks, CA: Sage.

Long, J. S., & Trivedi, P. K. 1993. Some specification tests for the linear regression model. *In*: Bollen, K. A., & Long, J. S. (eds), *Testing Structural Equation Models.* Newbury Park, CA: Sage.

Longford, N. T. 1987. A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested

random effects. *Biometrika*, 74, 817-827.

Longford, N. T. 1993. *Random Coefficient Models*. Oxford: Oxford University Press.

Longford, N. T. 1994. Logistic regression with random coefficients. *Computational Statistics & Data Analysis*, 17, 1-15.

Longford, N. T. 2001. Simulation-based diagnostics in random-coefficient models. *Journal of the Royal Statistical Society, Series A*, 164, 259-273.

Longford, N. T., & Muthén, B. O. 1992. Factor analysis for clustered observations. *Psychometrika*, 57, 581-597.

Lord, F. M. 1953. The relation of test score to the trait underlying the test. *Educational and Psychological Measurement*, 13, 517-549.

Lord, F. M. 1980. *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale, NJ: Erlbaum.

Lord, F. M., & Novick, M. R. 1968. *Statistical Theories of Mental Test Scores*. Reading: Addison-Wesley.

Louis, T. A. 1982. Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 44, 226-233.

Louis, T. A. 1984. Bayes and empirical Bayes estimates of a population of parameter values. *Journal of the American Statistical Association*, 79, 393-398.

Luce, R. D. 1959. *Individual Choice Behavior*. New York: Wiley.

Luce, R. D., & Suppes, P. 1965. Preference, utility and subjective probability. *In*: Luce, R., Bush, R., & Galanter, E. (eds), *Handbook of Mathematical Psychology III*. New York: Wiley.

Luijben, T. C. W. 1991. Equivalent models in covariance structure analysis. *Psychometrika*, 56, 653-665.

MacCallum, R. C., Wegener, D. T., Uchino, B. N., & Fabrigar, L. R. 1993. The problem of equivalent models in applications of covariance structure analysis. *Psychological Bulletin*, 114, 185-199.

MacDonald, I. L., & Zucchini, W. 1997. *Hidden Markov and Other Models for Discretevalued Time Series*. London: Chapman & Hall/CRC.

MacReady, G. B., & Dayton, C. M. 1992. The application of latent class models in adaptive testing. *Psychometrika*, 57, 71-88.

Maddala, G. S. 1971. The use of variance components models in pooling cross-section and time series data. *Econometrica*, 39, 341-358.

Maddala, G. S. 1977. *Econometrics*. New York: McGraw-Hill.

Maddala, G. S. 1983. *Limited Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.

Magder, S. M., & Zeger, S. L. 1996. A smooth nonparametric estimate of a mixing distribution using mixtures of Gaussians. *Journal of the American Statistical Association*, 11, 86-94.

Magidson, J., & Vermunt, J. K. 2002. Nontechnical introduction to latent class models. *Statistical Innovations White Paper* 1.

Manski, C. F. 1981. Structural models for discrete data: The analysis of discrete choice. *Pages* 58-109 *of*: Leinhardt, S. (ed), *Sociological Methodology* 1981. San Francisco, CA: Jossey-Bass.

Maple 9 Learning Guide. 2003. *Maple 9 Learning Guide*. Waterloo, ON: Maplesoft.

Marden, J. I. 1995. *Analyzing and Modeling Rank Data*. London: Chapman & Hall.

Maritz, J. S., & Lwin, T. 1989. *Empirical Bayes Methods*. London: Chapman & Hall.

Marsh, H. W., Balla, J. W., & McDonald, R. P. 1988. Goodness-of-fit indices in confirmatory factor analysis: Effects of sample size. *Psychological Bulletin*, 103, 391-411.

Marshall, E. C., & Spiegelhalter, D. J. 2003. Simulation-based tests for divergent behaviour in hierarchical models. Submitted for publication.

Martin, J. K., & McDonald, R. P. 1975. Bayesian estimation in unrestricted factor analysis: A treatment for Heywood cases. *Psychometrika*, 40, 505-517.

Marx, K. 1970. A note on classes. *Pages* 5-6 *of*: E. Laumann et al. (ed), *The Logic of Social Hierarchies*. Chicago, IL: Markham.

Mason, W. M., House, J. S., & Martin, S. S. 1985. On the dimensions of Political alienation in America. *Pages* 111-151 *of*: Tuma, N. B. (ed), *Sociological Methodology* 1985. San Francisco, CA: Jossey-Bass.

Masters, G. N. 1982. A Rasch model for partial credit scoring. *Psychometrika*, 47, 149-174.

Masters, G. N. 1985. A comparison of latent-trait and latent-class analyses of Likert-type data. *Psychometrika*, 50, 69-82.

Maxwell, A. E. 1971. Estimating true scores and their reliabilities in the case of composite psychological tests. *British Journal of Mathematical and Statistical Psychology*, 24, 195-204.

McCullagh, P. 1980. Regression models for ordinal data (with discussion). *Journal of the Royal Statistical Society, Series B*, 42, 109-142.

McCullagh, P. 1983. Quasi-likelihood functions. *The Annals of Statistics*, 11, 59-67.

McCullagh, P., & Nelder, J. A. 1989. *Generalized Linear Models (Second Edition)*. London: Chapman & Hall.

McCulloch, C. E. 1994. Maximum likelihood variance components estimation for binary data. *Journal of the American Statistical Association*, 89, 330-335.

McCulloch, C. E. 1997. Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American Statistical Association*, 92, 162-170.

McCulloch, C. E., & Searle, S. R. 2001. *Generalized, Linear and Mixed Models*. New York: Wiley.

McCulloch, C. E., Lin, H., Slate, E. H., & Turnbull, B. W. 2002. Discovering subpopulation structure with latent class

mixed models. *Statistics in Medicine*, 21, 417-429.

McCutcheon, A. L. 1987. *Latent Class Analysis*. Sage University Paper Series on Quantitative Applications in the Social Sciences. Beverly Hills, CA: Sage.

McDonald, R. P. 1967. *Nonlinear Factor Analysis*. Bowling Green, OH: Monograph 15, Psychometric Society.

McDonald, R. P. 1979. The simultanous estimation of factor loadings and scores. *British Journal of Mathematical and Statistical Psychology*, 32, 212-228.

McDonald, R. P. 1981. The dimensionality of tests and items. *British Journal of Mathematical and Statistical Psychology*, 34, 100-117.

McDonald, R. P. 1982. A note on the investigation of local and global identifiability. *Psychometrika*, 47, 101-103.

McDonald, R. P. 1985. *Factor Analysis and Related Methods*. Hillsdale, NJ: Erlbaum.

McDonald, R. P. 1989. An index of goodness-of-fit based on noncentrality. *Journal of Classification*, 6, 97-103.

McDonald, R. P., & Goldstein, H. 1989. Balanced and unbalanced designs for linear structural relations in two-level data. *British Journal of Mathematical and Statistical Psychology*, 42, 215-232.

McDonald, R. P., & Krane, W. R. 1977. A note on local identifiability and degrees of freedom in the asymptotic maximum likelihood test. *British Journal of Mathematical and Statistical Psychology*, 30, 198-203.

McFadden, D. L. 1973. Conditional logit analysis of qualitative choice behavior. *Pages* 105-142 *of*: Zarembka, P. (ed), *Frontiers in Econometrics*. New York: Academic Press.

McFadden, D. L. 1989. A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica*, 57, 995-1026.

McFadden, D. L., & Train, K. E. 2000. Mixed MNL models for discrete choice. *Journal of Applied Econometrics*, 15, 447-470.

McGilchrist, C. A. 1994. Estimation in generalized mixed models. *Journal of the Royal Statistical Society*, *Series B*, 56, 61-69.

McGilchrist, C. A., & Aisbett, C. W. 1991. Regression with frailty in survival analysis. *Biometrics*. 47, 461-466.

McGraw, K. O., & Wong, S. P. 1996. Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1, 30-46.

McKelvey, R., & Zavoina, W. 1975. A statistical model for the analysis of ordinal dependent variables. *Journal of Mathematical Sociology*, 4, 103-120.

McKendrick, A. G. 1926. Applications of mathematics to medical problems. *Proceedings of the Edinburgh Mathematical Society*, 44, 98-130.

McLachlan, G., & Basford, K. E. 1988. *Mixture Models. Inference and Applications to Clustering*. New York, NY: Dekker.

McLachlan, G., & Krishnan, T. 1997. *The EM Algorithm and Extensions*. Chichester: Wiley.

McLachlan, G., & Peel, D. A. 2000. *Finite Mixture Models*. New York, NY: Wiley.

Meijer, E., van der Leeden, R., & Busing, F. M. T. A. 1995. Implementing the bootstrap for multilevel models. *Multilevel Modelling Newsletter*, 7, 7-11.

Mendonça, L., & Böhning, D. 1994. Die Auswirking von Gesundheitsunterricht und Mundspühlung mit Na-Flourid auf die Prävention von Zahnkaries: eine Kohortenstudie mit urbanen Kindern in Brasilien. *In*: 39*th Conf. Deutsche Gesellschaft für Medizinische Informatik*, *Biometrie und Epidemiologie*.

Meng, X.-L., & Rubin, D. B. 1991. Using EM to obtain asymptotic variance-covariance matrices: The SEM Algorithm. *Journal of the American Statistical Association*, 86, 899-909.

Meng, X-L., & Rubin, D. B. 1992. Performing likelihood ratio tests with multiply-imputed data sets. *Biometrika*, 79, 103-111.

Meng, X-L., & Rubin, D. B. 1993. Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika*, 80, 267-278.

Meng, X-L., & Schilling, S. G. 1996. Fitting full-information item factor models and an empirical investigation of bridge sampling. *Journal of the American Statistical Association*, 91, 1254-1267.

Meredith, W. 1964. Notes on factorial invariance. *Psychometrika*, 29, 177-185.

Meredith, W. 1993. Measurement invariance, factor analysis and factor invariance. *Psychometrika*, 58, 525-544.

Meredith, W., & Tisak, J. 1990. Latent curve analysis. *Psychometrika*, 55, 107-122.

Messick, S. 1981. Constructs and their vicissitudes in educational and psychological measurement. *Psychological Bulletin*, 89, 575-588.

Miller, A. J. 1984. Selection of subsets of regression variables (with discussion). *Journal of the Royal Statistical Society*, *Series B*, 147, 389-425.

Miller, J. J. 1977. Asymptotic properties of maximum likelihood estimates in the mixed model of the analysis of variance. *The Annals of Statistics*, 5, 746-762.

Miller, W. E., Miller, A. H., & Schneider, E. J. 1980. *American National Election Studies Data Source-Book*: 1952-1978. Cambridge: Harvard University Press.

Mislevy, R. J. 1985. Estimation of latent group effects. *Journal of the American Statistical Association*, 80, 993-997.

Molenaar, P. C. W., & von Eye, A. 1994. On the arbitrary nature of latent variables. *Pages* 226-242 *of*: von Eye, A., & Clogg, C. C. (eds), *Latent Variable Analysis*: *Applications to Developmental Research*. Thousand Oaks, CA: Sage.

Molenberghs, G. 2002. Generalized estimating equations. *Pages* 47-75 *of*: Aerts, M., Geys, H., Molenberghs, G., & Ryan,

L. M. (eds), *Topics in Modelling Clustered Data*. Boca Raton, FL: Chapman & Hall/CRC.

Mollié, A. 1996. Bayesian mapping of disease. *Pages* 359-379 *of*: Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (eds), *Markov Chain Monte Carlo in Practice*. London: Chapman & Hall.

Moran, P. A. P. 1971. Maximum likelihood estimation in non-standard conditions. *Proceedings of the Cambridge Philosophical Society*, 70, 441-450.

Morris, C. 1983. Parametric empirical Bayes inference, theory and applications. *Journal of the American Statistical Association*, 78, 47-65.

Morris, J. N., Marr, J. W., & Clayton, D. G. 1977. Diet and heart: Postscript. *British Medical Journal*, 2, 1307-1314.

Moustaki, I. 1996. A latent trait and a latent class model for mixed observed variables. *British Journal of Mathematical and Statistical Psychology*, 49, 313-334.

Moustaki, I. 2000. A latent variable model for ordinal variables. *Applied Psychological Measurement*, 24, 211-223.

Moustaki, I., & Knott, M. 2000. Generalized latent trait models. *Psychometrika*, 65, 391-411.

Mukherjee, B. N. 1973. Analysis of covariance structures and exploratory factor analysis. *British Journal of Mathematical and Statistical Psychology*, 26, 125-154.

Mulaik, S. A. 1972. *The Foundations of Factor Analysis*. New York: McGraw-Hill.

Mulaik, S. A. 1988a. A brief history of the philosophical foundations of exploratory factor analysis. *Multivariate Behavioral Research*, 23, 267-305.

Mulaik, S. A. 1988b. Confirmatory factor analysis. *Pages* 259-288 *of*: Nesselroade, J. R., & Cattell, R. B. (eds), *Handbook of Multivariate Experimental Psychology*. New York: Plenum.

Mulaik, S. A., James, L. R., Alstine, J. Van, Bennett, N., Lind, S., & Stillwell, C. D. 1989. An evaluation of goodness of fit indices for structural equation models. *Psychological Bulletin*, 105, 430-445.

Müller, P., & Roeder, K. 1997. A Bayesian semiparametric model for case-control studies with errors in variables. *Biometrika*, 84, 523-537.

Muraki, E. 1990. Fitting a polytomous item response model to Likert-type data. *Applied Psychological Measurement*, 14, 59-71.

Muraki, E., & Bock, R. D. 1993. *PARSCALE*. Chicago, IL: Scientific Software.

Muraki, E., & Engelhard Jr., G. 1985. Full-information item factor analysis: Applications of EAP scores. *Applied Psychological Measurement*, 9, 417-430.

Muthén, B. O. 1977. *Statistical Methodology for Structural Equation Models involving Latent Variables with Dichotomous Indicators*. Ph. D. thesis, University of Uppsala.

Muthén, B. O. 1978. Contributions to factor analysis of dichotomous variables. *Psychometrika*, 43, 551-560.

Muthén, B. O. 1981. Factor analysis of dichotomous variables. *In*: Jackson, D. F., & Borgatta, E. F. (eds), *Factor Analysis and Measurement in Sociological Research: A Multidimensional Perspective*. London: Sage.

Muthén, B. O. 1982. Some categorical response models with continuous latent variables. *Pages* 65-79 *of*: Jöreskog, K. S., & Wold, H. (eds), *Systems under Indirect Observation-Causality ★ Structure ★ Prediction - Part II*. Amsterdam: North-Holland.

Muthén, B. O. 1983. Latent variable structural modelling with categorical data. *Journal of Econometrics*, 22, 43-65.

Muthén, B. O. 1984. A general structural equation model with dichotomous, ordered categorical and continuous latent indicators. *Psychometrika*, 49, 115-132.

Muthén, B. O. 1985. A method for studying the homogeneity of test items with respect to other relevant variables. *Journal of Educational Statistics*, 10, 121-132.

Muthén, B. O. 1988a. *LISCOMP: Analysis of Linear Structural Equations with a Comprehensive Measurement Model*. Mooresville, IN: Scientific Software.

Muthén, B. O. 1988b. Some uses of structural equation modelling in validity studies: Extending IRT to external variables. *Pages* 213-238 *of*: Wainer, H., & Braun, H. (eds), *Test Validity*. Hillsdale, NJ: Erlbaum.

Muthén, B. O. 1989a. Dichotomous factor analysis of symptom data. *Sociological Methods & Research*, 18, 19-65.

Muthén, B. O. 1989b. Multiple-group structural modelling with non-normal continuous variables. *British Journal of Mathematical and Statistical Psychology*, 42, 55-62.

Muthén, B. O. 1989c. Tobit factor analysis. *British Journal of Mathematical and Statistical Psychology*, 42, 241-250.

Muthén, B. O. 1989d. Using item-specific instructional information in achievement modeling. *Psychometrika*, 54, 385-396.

Muthén, B. O. 1993. Goodness of fit with categorical and other nonnormal variables. *Pages* 205-234 *of*: Bollen, K. A., & Long, J. S. (eds), *Testing Structural Equation Models*. Newbury Park, CA: Sage.

Muthén, B. O. 1994. Multilevel covariance structure analysis. *Sociological Methods & Research*, 22, 376-398.

Muthén, B. O. 2001. Second-generation structural equation modeling with a combination of categorical and continuous latent variables: New opportunities for latent class/latent growth modeling. *Pages* 291-322 *of*: Collins, L. M., & Sayer, A. (eds), *New Methods for the Analysis of Change*. Washington, DC: APA.

Muthén, B. O. 2002. Beyond SEM: General latent variable modeling. *Behaviormetrika*, 29, 81-117.

Muthén, B. O., & Kaplan, D. 1992. A comparison of some methodologies for the factoranalysis of nonnormal likert variables-a note on the size of the model. *British Journal of Mathematical and Statistical Psychology*, 45, 19-30.

Muthén, B. O., & Satorra, A. 1995. Complex sample data in structural equation modeling. *Pages* 267-316 *of*: Marsden, P. (ed), *Sociological Methodology* 1995. Cambridge, MA: Blackwell.

Muthén, B. O., & Satorra, A. 1996. Technical aspects of Muthén's LISCOMP approach to estimation of latent variable relations with a comprehensive measurement model. *Psychometrika*, 60, 489-503.

Muthén, B. O., & Shedden, K. 1999. Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics*, 55, 463-469.

Muthén, B. O., Du Toit, S. H. C., & Spisic, D. 1997. *Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes.* Unpublished paper.

Muthén, L. K., & Muthén, B. O. 1998. *Mplus User's Guide.* Los Angeles, CA: Muthén & Muthén.

Muthén, L. K., & Muthén, B. O. 2003. *Mplus Version* 2. 13. *Addendum to the Mplus User's Guide.* Los Angeles, CA: Muthén & Muthén. Downloadable from http ://www. statmodel. com/support/download.

Myles, J., & Clayton, D. G. 2001. *The GLMMGibbs package: Generalised linear mixed models by Gibbs sampling.* Downloadable from http://lib. stat. cmu. edu/R/CRAN/doc/packages/GLMMGibbs. pdf.

Nagin, D. S., & Land, K. C. 1993. Age, criminal careers, and population heterogeneity: Specification and estimation of a nonparametric mixed Poisson model. *Criminology*, 31, 327-362.

Natarajan, R., & Kass, R. E. 2000. Reference Bayesian methods for generalized linear mixed models. *Journal of the American Statistical Association*, 95, 227-237.

Naylor, J. C., & Smith, A. F. M. 1982. Applications of a method for the efficient computation of posterior distributions. *Journal of the Royal Statistical Society, Series C*, 31, 214-225.

Naylor, J. C., & Smith, A. F. M. 1988. Econometric illustrations of novel numerical integration strategies for Bayesian inference. *Journal of Econometrics*, 38, 103-125.

Neale, M. C., & Cardon, L. R. 1992. *Methodology for Genetic Studies of Twins and Families.* London: Kluwer.

Neale, M. C., Boker, S. M., Xie, G., & Maes, H. H. 2002. *Mx: Statistical Modeling (Sixth Edition).* Richmond, VA: Virginia Commonwealth University, Department of Psychiatry. Downloadable from http://www. vipbg. vcu. edu/mxgui/.

Nelder, J. A., & Wedderburn, R. W. M. 1972. Generalised linear models. *Journal of the Royal Statistical Society, Series A*, 135, 370-384.

Neuhaus, J. M. 1992. Statistical methods for longitudinal and clustered designs with binary responses. *Statistical Methods in Medical Research*, 1, 249-273.

Neuhaus, J. M., & Jewell, N. P. 1990. The effect of retrospective sampling on binary regression models for clustered data. *Biometrics*, 46, 977-990.

Neuhaus, J. M., Hauck, W. W., & Kalbfleisch, J. D. 1992. The effects of mixture distribution misspecification when fitting mixed-effects logistic models. *Biometrika*, 79, 755-762.

Neyman, J. 1939. On a new class of 'contagions' distributions, applicable in entomology and bacteriology. *Annals of Mathematical Statistics*, 10, 35-57.

Neyman, J., & Scott, E. L. 1948. Consistent estimates based on partially consistent observations. *Econometrica*, 16, 1-32.

Niemi, R. G., Craig, S. C., & Mattei, F. 1991. Measuring internal political efficacy in the 1988 National Election Study. *American Political Science Review*, 85, 1407-1413.

Normand, S.-L. T. 1999. Tutorial in biostatistics. Meta-analysis: Formulating, evaluating, combining, and reporting. *Statistics in Medicine*, 18, 321-359.

Novick, M. R., & Lewis, C. 1967. Coefficient alpha and the reliability of composite measurements. *Psychometrika*, 32, 1-13.

Novick, M. R., Lewis, C., & Jackson, P. H. 1973. The estimation of proportions in *M* groups. *Psychometrika*, 38, 19-46.

Nunnally, J. C., & Durham, R. L. 1975. Validity, reliability, and special problems of measurement in evaluation research. *In*: Struening, E. L., & Guttentag, M. (eds), *Handbook of Evaluation Research*, vol. 1. London: Sage.

Oakes, D. 1999. Direct calculation of the information matrix via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 61,479-482.

Oakes, M. 1993. The logic and role of meta-analysis in clinical research. *Statistical Methods in Medical Research*, 2, 147-160.

O'Hagan, A. 1976. On posterior joint and marginal modes. *Biometrika*, 63, 329-333.

Olsson, U. 1979. Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, 44, 443-460.

Olsson, U., Dragsow, F., & Dorans, N. J. 1982. The polyserial correlation coefficient. *Psychometrika*, 47, 337-347.

Orchard, T., & Woodbury, M. A. 1972. Missing information principle: Theory and applications. *Pages* 697-715 *of*: *Proceedings of the 6th Berkeley Symposium on Mathematical Statistics and Probability.* Berkeley, CA: University of California Press.

Otis, D. L., Burnham, K. P., White, G. C., & Anderson, D. R. 1978. Statistical inference from capture data on closed animal populations. *Wildlife Monographs*, 62, 1-135.

Ouwens, M. J. N. M., Tan, F. E. S., & Berger, M. P. R. 2001. Local influence to detect influential data structures for generalized linear mixed models. *Biometrics*, 57, 1166-1172.

Pan, J.-X., & Fang, K.-T. 2002. *Growth Curve Models and Statistical Diagnostics.* New York: Springer.

Pan, J. X., & Thompson, R. 2003. Gauss-Hermite quadrature approximation for estimation in generalised linear mixed models. *Computational Statistics*, 18, 57-78.

Parke, W. P. 1986. Pseudo maximum likelihood estimation: The asymptotic distribution. *The Annals of Statistics*, 14, 355-357.

Patterson, B. H., Dayton, C. M., & Graubard, B. I. 2002. Latent class analysis of complex sample survey data: Application to dietary data. *Journal of the American Statistical Association*, 97, 721-741.

Patterson, H. D., & Thompson, R. 1971. Recovery of inter-block information when block sizes are unequal. *Biometrika*, 58, 545-554.

Pawitan, Y. 2001. *In All Likelihood: Statistical Modelling and Inference Using Likelihood.* Oxford: Oxford University Press.

Payne, R. W. (ed). 2002. *The Guide to GenStat Release* 6. 1-*Part* 2: *Statistics.* Hemel Hempstead: VSN International.

Pearl, J. 2000. *Causality. Models, Reasoning, and Inference.* Cambridge: Cambridge University Press.

Pearson, K. 1901. Mathematical contributions to the theory of evolution VII. On the inheritance of characters not capable of exact quantitative measurement. *Philosophical Transactions of the Royal Society A*, 195, 79-150.

Pearson, K., & Heron, D. 1913. On theories of association. *Biometrika*, 9, 159-315.

Pendergast, J. F., Gange, S. J., Newton, M. A., Lindstrom, M. J., Palta, M., & Fisher, M. R. 1996. A survey of methods for analyzing clustered binary response data. *International Statistical Review.* 64. 89-118.

Peto, R. 1972. Contribution to the discussion of a paper by D. R. Cox. *Journal of the Royal Statistical Society, Series B*, 34, 205-207.

Pfeffermann, D. 2002. Small area estimation - new developments and directions. *International Statistical Review*, 70, 125-143.

Pickles, A. 1998. Generalized estimating equations. *Pages* 1626-1636 *of*: *Encyclopedia of Biostatistics.* London: Wiley.

Pickles, A., & Crouchley, R. 1994. Generalizations and applications of frailty models for survival and event data. *Statistical Methods in Medical Research*, 3, 263-278.

Pickles, A., & Crouchley, R. 1995. A comparison of frailty models for multivariate survival data. *Statistics in Medicine*, 14, 1447-1461.

Pickles, A., Pickering, K., & Taylor, C. 1996. Reconciling recalled dates of developmental milestones, events and transitions: a mixed generalized linear model with random mean and variance functions. *Journal of the Royal Statistical Society, Series A*, 159, 225-234.

Pickles, A., Pickering, K., Taylor, C., Sutton, C., & Yang, S. 2001. Multilevel risk models for retrospective age-of-onset data: School children's first cigarette. *Journal of Adolescent Research*, 16, 188-204.

Pinheiro, J. C., & Bates, D. M. 1995. Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational Graphics and Statistics*, 4, 12-35.

Pinheiro, J. C., & Bates, D. M. 2000. *Mixed-Effects Models in S and S-PLUS.* New York, NY: Springer.

Pinheiro, J. C., Liu, C., & Wu, Y. 2001. Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate *t*-distribution. *Journal of Computational Graphics and Statistics*, 10, 249-276.

Plackett, R. L. 1975. The analysis of permutations. *Journal of the Royal Statistical Society, Series C*, 24, 193-202.

Poon, W. -Y., & Lee, S. -Y. 1992. Maximum likelihood and generalized least squares analyses of two-level structural equation models. *Statistics and Probability Letters*, 14, 25-30.

Pregibon, D. 1981. Logistic regression diagnostics. *The Annals of Statistics*, 9, 705-724.

Prentice, R. L. 1988. Correlated binary regression with covariates specific to each binary observation. *Biometrics*, 44, 1033-1048.

Prentice, R. L. 1989. Surrogate endpoints in clinical trials: Definition and operational criteria. *Statistics in Medicine*, 8, 431-440.

Prentice, R. L., Williams, N. J., & Peterson, A. V. 1981. On the regression analysis of multivariate failure times. *Biometrika*, 68, 373-379.

Prochaska, J. O., & DiClemente, C. C. 1983 Stages and processes of self-change of smoking: Toward an integrative model of change. *Journal of Consulting and Clinical Psychology*, 51, 390-395.

Qu, Y., Tan, M., & Kutner, M. H. 1996. Random effects models in latent class analysis for evaluating accuracy of diagnostic tests. *Biometrics*, 52, 797-810.

Quandt, R. E. 1972. A new approach to estimating switching regressions. *Journal of the American Statistical Association*, 67, 306-310.

Rabe-Hesketh, S., & Everitt, B. S. 2003. *Handbook of Statistical Analyses using Stata (Third Edition).* Boca Raton, FL: Chapman & Hall/CRC.

Rabe-Hesketh, S., & Pickles, A. 1999. Generalised linear latent and mixed models. *Pages* 332-339 *of*: Friedl, H., Berghold, A., & Kauermann, C. (eds), 14*th International Workshop on Statistical Modeling.*

Rabe-Hesketh, S., & Skrondal, A. 2001. Parameterization of multivariate random effects models for categorical data. *Biometrics*, 57, 1256-1264.

Rabe-Hesketh, S., Pickles, A., & Taylor, C. 2000. sg129: Generalized linear latent and mixed models. *Stata Technical Bulletin*, 53, 47-57.

Rabe-Hesketh, S., Pickles, A., & Skrondal, A. 2001a. GLLAMM: A general class of multilevel models and a Stata program. *Multilevel Modelling Newsletter*, 13, 17-23.

Rabe-Hesketh, S., Pickles, A., & Skrondal, A. 2001b. *GLLAMM Manual.* Tech. rept. 2001/01. Department of Biostatistics and Computing, Institute of Psychiatry, King's College, University of London. Downloadable from http://www.gllamm.org/docum.html.

Rabe-Hesketh, S., Touloupulou, T., & Murray, R. M. 2001c. Multilevel modeling of cognitive function in schizophrenics

and their first degree relatives. *Multivariate Behavioral Research*, 36, 279-298.

Rabe-Hesketh, S., Yang, S., & Pickles, A. 2001d. Multilevel models for censored and latent responses. *Statistical Methods in Medical Research*, 10, 409-427.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. 2002. Reliable estimation of generalized linear mixed models using adaptive quadrature. *The Stata Journal*, 2, 1-21.

Rabe-Hesketh, S., Pickles, A., & Skrondal, A. 2003a. Correcting for covariate measurement error in logistic regression using nonparametric maximum likelihood estimation. *Statistical Modelling*, 3, 215-232.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A 2003b. Maxinmm likelihood estimation of generalized linear models with covariate measurement error. *The Stata Journal*, 3,386-411.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. 2004a. Generalized multilevel structural equation modeling. *Psychometrika*, 69, 167-190.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. 2004b. *GLLAMM Manual.* Tech. rept. 160. U. C. Berkeley Division of Biostatistics Working Paper Series. Downloadable from http://www. gllamm. org/docum. html.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. 2005a. Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *Journal of Econometrics*, in press.

Rabe-Hesketh, S., Pickles, A., & Skrondal, A. 2005b. *Multilevel and Structural Equation Modeling of Continuous, Categorical and Event Data.* College Station, TX: Stata Press.

Raftery, A. E. 1993. Bayesian model selection in structural equation models. *Pages* 163-180 *of*: Bollen, K. A., & Long, J. S. (eds), *Testing Structural Equation Models.* Newbury Park, CA: Sage.

Raftery, A. E. 1995. Bayesian model selection in social research. *Pages* 111-164 *of*: Raftery, A. E. (ed), *Sociological Methodology* 1995. Oxford: Blackwell.

Rao, C. R. 1975. Simultaneous estimation of parameters in different linear models and applications to biometric problems. *Biometrics*, 31, 545-554.

Rasbash, J., & Browne, W. J. 2001. Modelling non-hierarchical structures. *Pages* 91-106 *of*: Leyland, A. H., & Goldstein, H. (eds), *Multilevel Modelling of Health Statistics.* Chichester, UK: Wiley.

Rasbash, J., Steele, F., Browne, W., & Prosser, B. 2004. *A User's Guide to MLwiN Version* 2.0. London: Institute of Education. Downloadable from http://multilevel. ioe. ac. uk/download/manuals. html.

Rasch, G. 1960. *Probabilistic Models for Some Intelligence and Attainment Tests.* Copenhagen: Danmarks Pædagogiske Institut.

Rasch, G. 1961. On general laws and the meaning of measurement in psychology. *Pages* 321-333 *of*: *Proceedings of the IV Berkeley Symposium on Mathematical Statistics and Probability.* Berkeley, CA: University of California Press.

Rasch, G. 1967. An informal report on the theory of objectivity in comparisons. *Pages* 1-19 *of*: Van der Kamp, L. J. T., & Vlek, C. A. J. (eds), *Psychological Measurement Theory.* Leiden: University of Leiden.

Raudenbush, S. W., & Bryk, A. S. 2002. *Hierarchical Linear Models.* Thousand Oaks, CA: Sage.

Raudenbush, S. W., & Sampson, R. 1999a. Assessing direct and indirect effects in multilevel designs with latent variables. *Sociological Methods & Research*, 28, 123-153.

Raudenbush, S. W., & Sampson, R. 1999b. Ecometrics: Toward a science of assessing ecological settings, with application to the systematic social observation of neighborhoods. *Pages* 1-41 *of*: Marsden, P. V. (ed), *Sociological Methodology* 1999. Oxford: Blackwell.

Raudenbush, S. W., & Yang, M.-L. 1998. Numerical integration via high-order, multivariate LaPlace approximation with application to multilevel models. *Multilevel Modelling Newsletter*, 10, 11-14.

Raudenbush, S. W., Yang, M.-L., & Yosef, M. 2000. Maximum likelihood for generalized linear models with nested random effects via high-order, multivariate Laplace approximation. *Journal of Computational and Graphical Statistics*, 9, 141-157.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., Fai, Y., & Congdon, R. 2004. *HLM 6: Hierarchical Linear and Nonlinear Modeling.* Lincolnwood, IL: Scientific Software International.

Raykov, T., & Penev, S. 1999. On structural equation model equivalence. *Multivariate Behavioral Research*, 34, 199-244.

Reboussin, B. A., & Liang, K.-Y. 1998. An estimating equations approach for the LISCOMP model. *Psychometrika*, 63, 165-182.

Reiser, M. 1996. Analysis of residuals for the multinomial item response model. *Psychometrika*, 61, 509-528.

Reiser, M., & Lin, Y. 1999. A goodness-of-fit test for the latent class model when expected frequencies are small. *Pages* 81-111 *of*: Sobel M. E., & Becker, M. P. (eds), *Sociological Methodology* 1999. Cambridge: Blackwell.

Revelt, D., & Train, K. E. 1998. Mixed logit with repeated choices: Household's choices of appliance efficiency level. *Review of Economics and Statistics*, 80, 647-657.

Richardson, S., & Gilks, W. R. 1993. A Bayesian approach to measurement error problems in epidemiology using conditional independence models. *American Journal of Epidemiology*, 138, 430-442.

Richardson, S., Leblond, L., Jaussent, I., & Green, P. 2002. Mixture models in measurement error problems, with reference to epidemiological studies. *Journal of the Royal Statistical Society, Series A*, 165, 549-556.

Rijmen, F., Tuerlinckx, F., De Boeck, P., & Kuppens, P. 2003. A nonlinear mixed model framework for item response theory. *Psychological Methods*, 8, 185-205.

Rindskopf, D. 1992. A general approach to categorical data analysis with missing data, using generalized linear models with composite links. *Psychometrika*, 57, 29-42.

Rindskopf, D., & Rindskopf, W. 1986. The value of latent class analysis in medical diagnosis. *Statistics in Medicine*, 5, 21-27.

Ritz, J., & Spiegelman, D. 2004. A note about the equivalence of conditional and marginal regression models. *Statistical Methods in Medical Research*, 13, 309-323.

Rivers, D. 1988. Heterogeneity in models of electoral choice. *American Journal of Political Science*, 32, 737-757.

Roberts, J. S., Donoghue, J. R., & Laughlin, J. E. 2000. A general item response theory model for unfolding unidimensional polytomous responses. *Applied Psychological Measurement*, 24, 3-32.

Robins, J. M., Rotnitzky, A. G., & Zhao, L. P. 1994. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89, 826-866.

Robins, P. K., & West, R. W. 1977. Measurement errors in the estimation of home value. *Journal of the American Statistical Association*, 72, 290-294.

Robinson, G. K. 1991. That BLUP is a good thing: The estimation of random effects. *Statistical Science*, 6, 15-51.

Robinson, P. M. 1974. Identification, estimation, and large sample theory for regressions containing unobservable variables. *International Economic Review*, 15, 680-692.

Robinson, W. S. 1950. Ecological correlations and the behavior of individuals. *American Sociological Review*, 15, 351-357.

Rodriguez, G., & Goldman, N. 1995. An assessment of estimation procedures for multilevel models with binary responses: *Journal of the Royal Statistical Society, Series A*, 158, 73-89.

Rodriguez, G., & Goldman, N. 2001. Improved estimation procedures for multilevel models with binary response: A case study. *Journal of the Royal Statistical Society, Series A*, 164, 339-355.

Roeder, K., Carroll, R. J., & Lindsay, B. G. 1996. A semiparametric mixture approach to case-control studies with errors in covariables. *Journal of the American Statistical Association*, 91, 722-732.

Rosett, R. N., & Nelson, F. D. 1975. Estimation of a two-limit probit regression model. *Econometrica*, 43, 141-146.

Rosner, B., Spiegelman, D., & Willett, W. C. 1990. Correction of logistic regression relative risk estimates and confidence intervals for measurement error: The case of multiple covariates measured with error. *American Journal of Epidemiology*, 132, 734-745.

Rothenberg, T. 1971. Identification in parametric models. *Econometrica*, 39, 577-591.

Rotnitzky, A. G., & Wypij, D. 1994. A note on the bias of estimators with missing data. *Biometrics*, 50, 1163-1170.

Royston, P., & Altman, D. G. 1994. Regression using fractional polynomials of continuous covariates - parsimonious parametric modeling. *Journal of the Royal Statistical Society, Series C*, 43, 429-467.

Rubin, D. B. 1976. Inference and missing data. *Biometrika*, 63, 581-592.

Rubin, D. B. 1983. Some applications of Bayesian statistics to educational data. *Journal of the Royal Statistical Society, Series D*, 32, 155-167.

Rubin, D. B. 1984. Bayesianly justifiable and relevant frequency calculations for the applied statistician. *The Annals of Statistics*, 12, 1151-1172.

Rubin, D. B. 1987. *Multiple Imputation for Nonresponse in Surveys.* New York: Wiley.

Rubin, D. B. 1991. EM and beyond. *Psychometrika*, 56, 241-254.

Rubin, D. B., & Thayer, D. T. 1982. EM algorithms for ML factor analysis. *Psychometrika*, 47, 69-76.

Rubin, D. B., Stern, H. S., & Vehovar, V 1995. Handling "Don't Know" survey responses: The case of the Slovenian plebiscite. *Journal of the American Statistical Association*, 90, 822-828.

Rummel, R. J. 1967. *Applied Factor Analysis.* Evanston: Northwestern University Press.

Sackett, D. L., Haynes, R. B., Guyatt, G. H., & Tugwell, P. 1991. *Clinical Epidemiology.* Massachusetts: Little Brown & Company.

Samejima, F. 1969. *Estimation of Latent Trait Ability Using A Response Pattern of Graded Scores.* Bowling Green, OH: Psychometric Monograph 17, Psychometric Society.

Samejima, F. 1972. *A General Model for Free-Response Data.* Bowling Green, OH: Psychometric Monograph 18, Psychometric Society.

Sammel, M. D., Ryan, L. M., & Legler, J. M. 1997. Latent variable models for mixed discrete and continuous outcomes. *Journal of the Royal Statistical Society, Series B*, 59, 667-678.

Sanathanan, L. 1972. Estimating the size of a multinomial population. *The Annals of Mathematical Statistics*, 43, 142-152.

SAS/Stat User's Guide, version 8. 2000. *SAS/Stat User's Guide, version* 8. Cary, NC: SAS Publishing.

Satorra, A. 1990. Robustness issues in structural equation modeling: A review of recent developments. *Quality and Quantity*, 24, 367-386.

Satorra, A. 1992. Asymptotic robust inferences in the analysis of mean and covariance structures. *Pages* 249-278 *of*: Marsden, P. V. (ed), *Sociological Methodology* 1992. Washington, DC: American Sociological Association.

Satorra, A., & Bentler, P. M. 1994. Corrections to test statistics and standard errors in covariance structure analysis. *Pages* 285-305 *of*: von Eye, A., & Clogg, C. C. (eds), *Latent Variable Analysis: Applications to Developmental Research.* Newbury Park, CA: Sage.

Satten, G. A., & Kupper, L. L. 1993. Inferences about exposure-disease association using probability-of-exposure information. *Journal of the American Statistical Association*, 88, 200-208.

Schafer, D. W. 2001. Semiparametric maximum likelihood for measurement error regression. *Biometrics*, 57, 53-61.

Schafer, J. L. 1997. *Analysis of Incomplete Multivariate Data.* London: Chapman & Hall.

Schall, R. 1991. Estimation in generalized linear-models with random effects. *Biometrika*, 78, 719-727.

Schlesselman, J. J. 1982. *Case-Control Studies: Design, Conduct, Analysis.* Oxford: Oxford University Press.

Schluchter, M. D. 1988. Analysis of incomplete multivariate data using linear models with structured covariance matrices. *Statistics in Medicine*, 7, 317-324.

Schoenberg, R. 1985. Latent variables in the analysis of limited dependent variables. *Pages* 213-242 *of*: Tuma, N. B. (ed), *Sociologicai Methodology* 1985. San Francisco, CA: Jossey-Bass.

Schoenberg, R. 1996. *Constrained Maximum Likelihood.* Maple Valley, WA: Aptech Systems and The University of Washington Downloadable from http://faculty. washington. edu/rons/.

Schoenberg, R., & Richtand, C. 1984. An application of the EM method to the maximum likelihood estimation of multiple indicator and factor analysis models. *Sociological Methods & Research*, 13, 127-150.

Schumaker, R. E., & Marcoulides, G. A. (eds). 1998. *Interaction and Nonlinear Effects in Structural Equation Modeling.* Mahwah, NJ: Erlbaum.

Schwarz, G. 1978. Estimating the dimension of a model. *The Annals of Statistics*, 6, 461-464.

Searle, S. R., Casella, G., & McCulloch, C. E. 1992. *Variance Components.* New York: Wiley.

Self, S. G., & Liang, K.-Y. 1987. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under non-standard conditions. *Journal of the American Statistical Association*, 82, 605-610.

Seong, T.-J. 1990. Sensitivity of marginal maximum likelihood estimation of item and ability parameters to the characteristics of the prior ability distributions. *Applied Psychological Measurement*, 14, 299-311.

Serfling, R. J. 1980. *Approximation Theorems of Mathematical Statistics.* New York: Wiley.

Shavelson, R. J., & Webb, N. M. 1991. *Generalizability Theory: A Primer.* Newbury Park, CA: Sage.

Shaw, J. E. H. 1988. A quasirandom approach to integration in Bayesian statistics. *The Annals of Statistics*, 16, 895-914.

Shepard, R. N. 1974. Representation of structure in similarity data: Problems and prospects. *Psychometrika*, 39, 373-421.

Shrout, P. E., & Fleiss, J. L. 1979. Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86, 420-428.

Sijtsma, K., & Molenaar, I. W. 2002. *Introduction to Nonparametric Item Response Theory.* Thousand Oaks, CA: Sage.

Silagy, C. 2003. Nicotine replacement therapy for smoking cessation (Cochrane review). *In*: *The Cochrane Library*, *Issue* 4. Chichester: Wiley.

Silva, F. 1993. *Psychometric Foundations and Behavioral Assessment.* Newbury Park, CA: Sage.

Simar, L. 1976. Maximum likelihood estimation of a compound Poisson process. *The Annals of Statistics*, 4, 1200-1209.

Singer, J. D., & Willett, J. B. 1993. It's about time: Using discrete-time survival analysis to study duration and the timing of events. *Journal of Educational Statistics*, 18, 155-195.

Singer, J. D., & Willett, J. B. 2003. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence.* New York: Oxford University Press.

Skaug, H. J. 2002. Automatic differentiation to facilitate maximum likelihood estimation in nonlinear random effects models. *Journal of Computational & Graphical Statistics*, 11, 458-470.

Skellam, J. G. 1948. A probability distribution derived from the binomial distribution by regarding the probability of a success as variable between the sets of trials. *Journal of the Royal Statistical Society*, *Series B*, 10, 257-261.

Skrondal, A. 1996. *Latent Trait, Multilevel and Repeated Measurement Modelling with Incomplete Data of Mixed Measurement Levels.* Oslo: Section of Medical Statistics, University of Oslo.

Skrondal A. 2000. Design and analysis of Monte Carlo experiments: Attacking the conventional wisdom. *Multivariate Behavioral Research*, 35, 137-167.

Skrondal A. 2003. Interaction as departure from additivity in case-control studies: A cautionary note. *American Journal of Epidemiology*, 158, 251-258.

Skrondal A., & Laake, P. 1999. Factorial invariance in confirmatory factor analysis. *Technical Report.*

Skrondal A., & Laake, P. 2001. Regression among factor scores. *Psychometrika*, 66, 563-576.

Skrondal A., & Rabe-Hesketh, S. 2003a. Generalized linear mixed models for nominal data. *American Statistical Association, Proceedings of the Biometrics Section*, 3931-3936.

Skrondal A., & Rabe-Hesketh, S. 2003b. Multilevel logistic regression for polytomous data and rankings. *Psychometrika*, 68, 267-287.

Skrondal A., & Rabe-Hesketh, S. 2003c. Some applications of generalized linear latent and mixed models in epidemiology: Repeated measures, measurement error and multilevel modelling. *Norwegian Journal of Epidemiology*, 13, 265-278.

Skrondal, A., & Rabe-Hesketh, S. 2004. Generalised linear latent and mixed models with composite links and exploded likelihoods. *Pages* 27-39 *of*: A. Biggeri, E. Dreassi, C. Lagazio, & M. Marchi (eds), 19*th International Workshop on Statistical Modeling.* Florence, Italy: Firenze University Press.

Skrondal, A., Eskild, A., & Thorvaldsen, J. 2000. Changes in condom use after HIV diagnosis. *Scandinavian Journal of Public Health*, 28, 71-76.

Skrondal, A., Rabe-Hesketh, S., & Pickles, A. 2002. Informative dropout and measurement error in cluster randomised trials. *Page* 61 *of*: *Proceedings of the XXIth International Biometric Conference (Abstracts).*

Smith, A. F. M. 1973. A general Bayesian linear model. *Journal of the Royal Statistical Society*, *Series B*, 35, 61-75.

Smith, T. W. 1988. *Rotation Design of the GSS.* Tech. rept. 52. GSS Methodological Report, Chicago.

Snijders, T. A. B., & Berkhof, J. 2004. Diagnostic checks for multilevel models. *Page Chapter* 4 *of*: de Leeuw, J, & Kreft,

I. (eds), *Handbook of Quantitative Multilevel Analysis.* Thousand Oaks, CA: Sage.

Snijders, T. A. B., & Bosker, R. J. 1999. *Multilevel Analysis.* London: Sage.

Sobel, M. E. 1995. Causal inference in the social and behavioral sciences. *Pages* 1-38 *of*: Arminger, G., Clogg, C. C., & Sobel, M. E. (eds), *Handbook of Statistical Modelling for the Social and Behavioral Sciences.* New York: Plenum Press.

Sobel, M. E., & Bohrnstedt, G. W. 1985. Use of null models in evaluating the fit of covariance structure models. *Pages* 152-178 *of*: Tuma, N. B. (ed), *Sociological Methodology* 1985. San Francisco, CA: Jossey-Bass.

Social and Community Planning Research. 1987. *British Social Attitudes Panel Survey*, 1983-1986 [*Computer file*] *SN*: 2197. Colchester, Essex: The Data Archive [Distributor].

Sommer, A., Katz, J., & Tarwotjo, I. 1983. Increased mortality in children with slight vitamin A deficiency. *American Journal of Clinical Nutrition*, 40, 1090-1095.

Spearman, C. 1904. General intelligence, objectively determined and measured. *American Journal of Psychology*, 15, 201-293.

Spearman, C. 1927. *The Abilities of Man: Their Nature and Measurement.* London: Macmillan.

Spiegelhalter, D. J., Freedman, L. S., & Parmar, M. K. B. 1994. Bayesian approaches to randomized trials (with discussion). *Journal of the Royal Statistical Society*, *Series A*, 157, 357-416.

Spiegelhalter, D. J., Thomas, A., Best, N. G., & Gilks, W. R. 1996a. *BUGS* 0.5 *Bayesian Analysis using Gibbs Sampling. Manual* (*version ii*). Cambridge: MRC-Biostatistics Unit. Downloadable from http://www.mrcbsu.cam.ac.uk/bugs/documentation/contents.shtml.

Spiegelhalter, D. J., Thomas, A., Best, N. G., & Gilks, W. R. 1996b. *BUGS* 0.5 *Examples*, *Volume* 1. Cambridge: MRC-Biostatistics Unit.

Spiegelhalter, D. J., Thomas, A., Best, N. G., & Gilks, W. R. 1996c. *BUGS* 0.5 *Examples*, *Volume* 2. Cambridge: MRC-Biostatistics Unit.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. 2002. Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society*, *Series B*, 64, 583-639.

SPSS Inc. 2001. *SPSS* 11.0 *Advanced Models.* Prentice Hall: Englewood Cliffs, NJ.

Stata Cross-Sectional Time-Series. 2003. *Cross-Sectional Time-Series Reference Manual.* College Station, TX: Stata Press.

StataCorp. 2003. *Stata Statistical Software*: *Release* 8.0. College Station, TX.

Steiger, J. H. 1990. Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, 25, 173-180.

Stelzl, I. 1986. Changing causal relationships without changing the fit: Some rules for generating equivalent path models. *Multivariate Behavioral Research*, 21,309-331.

Sterlin, T. D. 1959. Publication decisions and their possible effects on inferences drawn from tests of significance - or vice-versa. *Journal of the American Statistical Association*, 54, 30-34.

Stern, S. 1992. A method for smoothing simulated moments of discrete probabilities in multinomial probit models. *Econometrica*, 60, 943-952.

Stern, S. 1997. Simulation-based estimation. *Journal of Economic Literature*, 35, 2006-2039.

Stevens, S. S. 1951. Mathematics, measurement and psychophysics. *In*: Stevens, S. S. (ed), *Handbook of Experimental Psychology.* New York: Wiley.

Stewart, M. B. 1983. On the least squares estimation when the dependent variable is grouped. *Review of Economic Studies*, 50, 737-753.

Stiratelli, R., Laird, N. M., & Ware, J. H. 1984. Random effects models for serial observations with binary responses. *Biometrics*, 40, 961-971.

Stone, M. 1974. Cross-validatory choice and assessment of statistical predictions (with discussion). *Journal of the Royal Statistical Society*, *Series B*, 36, 111-147.

Stram, D. O., & Lee, J. W. 1994. Variance components testing in the longitudinal mixed effects model. *Biometrics*, 50, 1171-1177.

Stram, D. O., & Lee, J. W. 1995. Correction to 'Variance components testing in the longitudinal mixed effects model'. *Biometrics*, 51, 1196.

Streiner, D. L., & Norman, G. R. 1995. *Health Measurement Scales*: *A Practical Guide to Their Development and Use.* Oxford: Oxford University Press.

Strenio, J. L. F., Weisberg, H. I., & Bryk, A. S. 1983. Empirical Bayes estimation of individual growth curve parameters and their relations to covariates. *Biometrics*, 39, 71-86.

Stroud, A. H. 1971. *Approximate Calculation of Multiple Integrals.* Englewood Cliffs, NJ: Prentice-Hall.

Stroud, A. H., & Secrest, D. 1966. *Gaussian Quadrature Fornulas.* Englewood Cliffs, NJ: Prentice-Hall,

Sudman, S., & Bradburn, N. M. 1973. Effects of time and memory factor on response in surveys. *Journal of the American Statistical Association*, 63, 805-815.

Sutton, A. J., Song, F., Gilbody, S. M., & Abrams, K. R. 2000. Modelling publication bias in meta-analysis: A review. *Statistical Methods in Medical Research*, 9, 421-445.

Swaminathan, H., & Gifford, J. A. 1982. Bayesian estimation in the Rasch model. *Journal of Educational Statistics*, 7, 175-192.

Swaminathan, H., & Gifford, J. A. 1985. Bayesian estimation in the two-parameter logistic model. *Psychometrika*, 50, 349-

364.
Swaminathan, H., & Gifford, J. A. 1986. Bayesian estimation in the three-parameter logistic model. *Psychometrika*, 51, 589-601.
Takane, Y. 1987. Analysis of covariance structures and probabilistic binary data. *Communication & Cognition*, 20, 45-62.
Takane, Y., & de Leeuw, J. 1987. On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika*, 52, 393-408.
Tanaka, J. S. 1993. Multifaceted conceptions of fit in structural equation models. *Pages* 10-39 *of*: Bollen, K. A., & Long, J. S. (eds), *Testing Structural Equation Models*. Newbury Park, CA: Sage.
Tanaka, J. S., Panter, A. T., Winborne, W. C., & Huba, G. J. 1990. Theory testing in personality and personality psychology with structural equation models. A primer in twenty questions. *Pages* 217-242 *of*: Hendrick, C., & Clark, M. S. (eds), *Review of Personality and Social Psychology*. Newbury Park, CA: Sage.
Tanner, M. A. 1996. *Tools for Statistical Inference (Third Edition)*. New York: Springer.
Tanner, M. A., & Wong, W. H. 1987. The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82, 528-540.
Ten Berge, J. M. F. 1983. On Green best linear composites with a specified structure, and oblique estimates of factor scores. *Psychometrika*, 48, 371-375.
Ten Berge, J. M. F., Krijnen, W. P., Wansbeek, T. J., & Shapiro, A. 1999. Some new results on correlation-preserving factor scores prediction methods. *Linear Algebra And Its Applications*, 289, 311-318.
Ten Have, T. R., & Localio, A. R. 1999. Empirical Bayes estimation of random effects parameters in mixed effects logistic regression models. *Biometrics*, 55, 1022-1029.
Terza, J. V. 1985. Ordinal probit: A generalization. *Communications in Statistics, Theory and Methods*, 14, 1-11.
Terza, J. V. 1998. Estimating count data models with endogenous switching: Sample selection and endogenous treatment effects. *Journal of Econometrics*, 84, 129-154.
Thall, P. F., & Vail, S. C. 1990. Some covariance models for longitudinal count data with overdispersion. *Biometrics*, 46, 657-671.
Therneau, T. M., & Grambsch, P. M. 2000. *Modeling Survival Data*. New York: Springer.
Thissen, D. 1982. Marginal maximum likelihood estimation for the one-parameter logistic model. *Psychometrika*, 47, 175-186.
Thissen, D., & Steinberg, L. 1986. A taxonomy of item response models. *Psychometrika*, 51, 567-577.
Thissen, D., & Steinberg, L. 1988. Data analysis using item response theory. *Psychological Bulletin*, 104, 385-395.
Thisted, R. A. 1987. *Elements of Statistical Computing*. London: Chapman and Hall.
Thompson, B. 1984. *Canonical Correlation Analysis: Uses and Interpretation*. Sage University Paper Series on Quantitative Applications in the Social Sciences. Beverly Hills, CA: Sage.
Thompson, R., & Baker, R. J. 1981. Composite link functions in generalized linear models. *Journal of the Royal Statistical Society, Series C*, 30, 125-131.
Thompson, S. G., & Pocock, S. J. 1991. Can meta-analysis be trusted? *The Lancet*, 338, 1127-1130.
Thompson, S. G., Turner, R. M., & Warn, D. E. 2001. Multilevel models for meta-analysis, and their application to absolute risk differences. *Statistical Methods in Medical Research*, 10, 375-392.
Thompson, W. A. 1977. On the treatment of grouped observations in life studies. *Biometrics*, 33, 463-470.
Thomson, G. H. 1938. *The Factorial Analysis of Human Ability*. London: University of London Press.
Thoresen, M., & Laake, P. 2000. A simulation study of measurement error correction methods in logistic regression. *Biometrics*, 56, 868-872.
Thurstone. L. L. 1935. *The Vectors of Mind*. Chicago: University of Chicago Press.
Thurstone, L. L. 1947. *Multiple-Factor Analysis: A Development And Expansion of the Vectors of Mind*. Chicago: University of Chicago Press.
Tierney, L., & Kadane, J. B. 1986. Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, 81, 82-86.
Titterington, D. M., Smith, A. F. M., & Makov, U. E. 1985. *Statistical Analysis of Finite Mixture Distributions*. Chichester: Wiley.
Tobin, J. 1958. Estimation of relationships for limited dependent variables. *Econometrica*, 26, 24-36.
Torgerson, W. S. 1958. *Theory and Methods of Scaling*. New York: Wiley.
Træen, B. 2003. Effect of an intervention to prevent unwanted pregnancy in adolescents. A randomised, prospective study from Nordland County, Norway, 1999-2001. *Journal of Community & Applied Social Psychology*, 13, 224-239.
Train, K. E. 1986. *Qualitative Choice Analysis*. Cambridge, MA: MIT Press.
Train, K. E. 2002. *Mixed Logits with Bounded Distributions*. Berkeley, CA: Department of Economics, University of California Berkeley. Downloadable from http ://elsa. berkeley. edu/ ~ train/software. html.
Train, K. E. 2003. *Discrete Choice Methods with Simulation*. Cambridge: Cambridge University Press.
Train, K. E., Revelt, D., & Ruud, P. A. 1999. *Mixed Logit Estimation Routine for Panel Data*. Berkeley, CA: Department of Economics, University of California Berkeley. Downloadable from http ://elsa. berkeley. edu/ ~ train/software. html.
Tucker, L. R. 1972. Relations between multidimensional scaling and three-mode factor analysis. *Psychometrika*, 37, 3-27.
Turner, R. M., Omar, R. Z., Yang, M., Goldstein, H., & Thompson, S. G. 2000. A multilevel framework for meta-

analysis of clinical trials with binary outcomes. *Statistics in Medicine*, 19, 3417-3432.

U. S. Department of Defense. 1982. *Profile of American Youth*. Washington DC: Office of the Assistant Secretary of Defense for Manpower, Reserve Affairs, and Logistics.

Uebersax, J. S. 1993. Statistical modeling of expert ratings on medical treatment appropriateness. *Journal of the American Statistical Association*, 88, 421-427.

Uebersax, J. S., & Grove, W. M. 1993. A latent trait finite mixture model for the analysis of rating agreement. *Biometrics*, 49, 823-835.

Vaida, F., & Blanchard, S. 2004. Conditional Akaike information for mixed effects models. *Submitted for publication.*

Vaida, F., & Xu, R. 2000. Proportional hazards model with random effects. *Statistics in Medicine*, 19, 3309-3324.

van de Pol, F., & Langeheine, R. 1990. Mixed Markov latent class models. *Pages* 213-247 *of*: Clogg, C. C. (ed), *Sociological Methodology* 1990. Oxford: Blackwell.

van den Berg, G. J. 2001. Duration models: specification, identification, and multiple durations. *In*: Heckman, J. J., & Learner, E. (eds), *Handbook of Econometrics*, vol. V. Amsterdam: North-Holland.

van der Leeden, R., & Busing, F. M. T. A. 1994. *First iteration versus IGLS/RIGLS estimates in two-level models: A Monte Carlo study with ML3. Preprint PRM* 94-03. Tech. rept. Psychometrics and Research Methodology, Leiden, The Netherlands.

van der Linden, W. J., & Hambleton, R. K. 1997. *Handbook of Modern Item Response Theory.* New York, NY: Springer.

van Houwelingen, J. C., Arends, L. R., & Stijnen, T. 2002. Tutorial in biostatistics. Advanced methods in meta-analysis: multivariate approach and meta-regression. *Statistics in Medicine*, 21, 589-624.

Vassend, O., & Skrondal, A. 1995. Factor-analytic studies of the NEO personality-inventory and the five-factor model - the problem of high structural complexity and conceptual indeterminacy. *Personality and Individual Differences*, 19, 135-147.

Vassend, O., & Skrondal, A. 1997. Validation of the NEO personality inventory and the five-factor model. Can findings from exploratory and confirmatory factor analysis be reconciled? *European Journal of Personality*, 11, 147-166.

Vassend, O., & Skrondal, A. 1999. The problem of structural indeterminacy in multidimensional symptom report instruments. The case of SCL-90-R. *Behaviour Research and Therapy*, 37, 685-701.

Vassend, O., & Skrondal, A. 2004. The five-factor model of personality. Language, factor structure, and the problem of generalization. *Submitted for publication.*

Vaupel, J. W., Manton, K. G., & Stallard, E. 1979. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, 16, 439-454.

Venables, W. N., & Ripley, B. D. 2002. *Modern Applied Statistics with S (Fourth Edition)*. New York, NY: Springer.

Verbeke, G., & Lesaffre, E. 1996. A linear mixed-effects model with heterogeneity in the random-effects population. *Journal of the American Statistical Association*, 91,217-221.

Verbeke, G., & Molenberghs, G. 1997. *Linear Mixed Models in Practice : A SAS-Oriented Approach.* New York, NY: Springer.

Verbeke, G., & Molenberghs, G. 2000. *Linear Mixed Models for Longitudinal Data.* New York, NY: Springer.

Verbeke, G., & Molenberghs, G. 2003. The use of score tests for inference on variance components. *Biometrics*, 59, 254-262.

Verbyla, A. P. 1993. Modeling variance heterogeneity: residual maximum likelihood and diagnostics. *Journal of the Royal Statistical Society, Series B*, 55, 493-508.

Vermunt, J. K. 1997. *Log-Linear Models for Event Histories*. Thousand Oaks, CA: Sage.

Vermunt, J. K. 2001. The use of restricted latent class models for defining and testing nonparametric item response theory models. *Applied Psychological Measurement*, 25,283-294.

Vermunt, J. K. 2003. Multilevel latent class models. *Pages* 213-239 *of*: Stolzenberg, R. M. (ed), *Sociological Methodology* 2003, vol. 33. Oxford: Blackwell.

Vermunt, J. K. 2004. An EM algorithm for the estimation of parametric and nonparametric hierarchical nonlinear models. *Statistiea Neerlandica*, 58, 220-233.

Vermunt, J. K., & Magidson, J. 2000. *Latent Gold User's Guide.* Belmont, MA: Statistical Innovations.

Vermunt, J. K., & Magidson, J. 2003a. *Addendum to the Latent GOLD Users Guide: Upgrade Manual for Version* 3. 01. Belmont, MA: Statistical Innovations.

Vermunt, J. K., & Magidson, J. 2003b. Latent class models for classification. *Computational Statistics & Data Analysis*, 41, 531-537.

Vinokur, A. D., Price, R. H., & Schul, Y. 1995. Impact of JOBS intervention on unemployed workers varying in risk for depression. *American Journal of Community Psychology*, 19, 543-562.

Vonesh, E. F., & Chinchilli, V. M. 1997. *Linear and Nonlinear Models for the Analysis of Repeated Measurements.* New York: Marcel Dekker.

Waclawiw, M. A., & Liang, K.-Y. 1994. Empirical Bayes estimation and inference for the random effects model with binary response. *Statistics in Medicine*, 13, 541-551.

Wainer, H. 1976. Estimating coefficients in linear models: It don't make no nevermind. *Psychological Bulletin*, 83, 213-217.

Wainer, H., & Kiely, G. L. 1987. Item clusters and computerized adaptive testing: A case for testlets. *Journal of Educational Measurement*, 24, 185-201.

Wainer, H., & Thissen, D. 1982. Some standard errors in item response theory. *Psychometrika*, 47, 397-412.

Wald, A. 1950. Note on the identification of economic relations. *In*: Koopmans, T. C. (ed), *Statistical Inference in Dynamic Economic Models*. New York: Cowles Commission Monograph 10, Wiley.

Wang, M. W., & Stanley, J. C. 1970. Differential weighting: A review of methods and empirical studies. *Review of Educational Research*, 40,663-705.

Wansbeek, T. J., & Meijer, E. (eds). 2002. *Measurement Error and Latent Variables in Econometrics*. Amsterdam: North Holland.

Warm, T. A. 1989. Weighted likelihood estimation of ability in item response models. *Psychometrika*, 54, 427-450.

Wasserman, L. 2000. Bayesian model selection and model averaging. *Journal of Mathematical Psychology*, 44, 92-107.

Waternaux, C., Laird, N. M., & Ware, J. H. 1989. Methods for analysis of longitudinal data: bloodlead concentrations and cognitive development. *Journal of the American Statistical Association*, 84, 33-41.

Wedderburn, R. W. M. 1974. Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika*, 61, 439-447.

Wedel, M. 2002a. Concomitant variables in finite mixture models. *Statistica Neerlandica*, 56, 362-375.

Wedel, M. 2002b. *GLIMMIX. A Program for Estimation of Latent Class Mixture and Mixture Regression Models. User's Guide Manual*. Groningen: ProGamma.

Wedel, M., & Kamakura, W. A. 2000. *Market Segmentation: Conceptual and Methodological Foundations (Second Edition)*. Dordrecht: Kluwer.

Wedel, M., & Kamakura, W. A. 2001. Factor analysis with (mixed) observed and latent variables. *Psychometrika*, 66, 515-530.

Wedel, M., Böckenholt, U., & Kamakura, W. A. 2003. Factor models for multivariate count data. *Journal of Multivariate Analysis*, 87, 356-369.

Wei, G. C. G., & Tanner, M. A. 1990. A Monte Carlo implementation of the EM algorithm and the poor man's augmentation algorithms. *Journal of the American Statistical Association*, 86, 669-704.

Wei, L. J., Lin, D. Y., & Weissfeld, L. 1989. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association*, 84, 1065-1073.

Werts, C. E., Linn, R. L., & Jöreskog, K. G. 1974. Intraclass reliability estimates: Testing structural assumptions. *Educational and Psychological Measurement*, 34, 25-33.

Wheaton, B., Muthén, B. O., Alwin, D. F., & Summers, G. 1977. Assessing reliability and stability in panel models. *In*: Heise, D. R. (ed), *Sociological Methodology* 1977. San Francisco, CA: Jossey-Bass.

White, H. 1982. Maximum likelihood estimation of misspecified models. *Econometrica*, 50,1-25.

Whitehead, A. 2002. *Meta-analysis of Controlled Clinical Trials*. Chichester: Wiley.

Whitehead, J. 1980. Fitting Cox's regression model to survival data using GLIM. *Journal of the Royal Statistical Society, Series C*, 29, 269-275.

Wieken-Mayser, M. 1979. *Political Action: An Eight Nation Study* 1973-1976. Köln: Zen-tralarchiv für Empirische Sozialforschung.

Wiggins, R. D., Ashworth, K., O'Muircheartaigh, C. A., & Galbraith, J. I. 1990. Multilevel analysis of attitudes to abortion. *Journal of the Royal Statistical Society, Series D*, 40, 225-234.

Wiley, D. E. 1973. The identification problem for structural equation models with unmeasured variables. *Pages* 69-83 *of*: Goldberger, A. S., & Duncan, O. D. (eds), *Structural Equation Models in the Social Sciences*. New York: Academic Press.

Wilks, S. S. 1938. Weighting systems for linear functions of correlated variables when there is no dependent variable. *Psychometrika*, 3, 23-40.

Williams, D. A. 1975. The analysis of binary responses from toxological experiments involving reproduction and teratogenecity. *Journal of the Royal Statistical Society, Series C*, 31, 949-952.

Williams, R. L. 2000. A note on robust variance estimation for cluster-correlated data. *Biometrics*, 56, 645-646.

Wilson, M., & Adams, R. J. 1995. Rasch models for item bundles. *Psychometrika*, 60, 181-198.

Winkelmann, R. 2003. *Econometric Analysis of Count Data*. New York: Springer.

Wolfe, R., & Firth, D. 2002. Modelling subjective use of an ordinal response scale in a many period crossover experiment. *Journal of the Royal Statistical Society, Series C*, 51, 245-255.

Wolfinger, R. D. 1999. *Fitting Non-Linear Mixed Models with the New NLMIXED procedure*. Tech. rept. SAS Institute, Cary, NC.

Wolfinger, R. D., & O'Connell, M. 1993. Generalized linear mixed models: A pseudolikelihood approach. *Journal of Statistical Computing and Simulation*, 48, 233-243.

Wolfram, S. 2003. *The Mathematica Book 5 (Fifth Edition)*. Champaign, IL: Wolfram Media.

Wooldridge, J. M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: The MIT Press.

Woolf, B. 1955. On estimating the relationship between blood groups and disease. *Annals of Human Genetics*, 19, 251-253.

Wright, B. D. 1977. Solving measurement problems with the Rasch model. *Journal of Educational Measurement*, 14, 97-116.

Wu, M. C., & Carroll, R. J. 1988. Estimation and comparison of change in the presence of informative right censoring by modeling the censoring process. *Biometrics*, 44, 175-188.

Wulfsohn, M. S., & Tsiatis, A. A. 1997. A joint model for survival and longitudinal data measured with error. *Biometrics*, 53, 330-339.

Xiang, L., Tse, S.-K., & Lee, A. H. 2002. Influence diagnostics for generalized linear mixed models: Applications to clustered data. *Computational Statistics & Data Analysis*, 40, 759-774.

Xu, J., & Zeger, S. L. 2001. Joint analysis of longitudinal data comprising repeated measures and times to events. *Journal of the Royal Statistical Society, Series C*, 50, 375-387.

Yang, M. 2001. Multinomial regression. *Pages* 107-123 *of*: Leyland, A. H., & Goldstein, H. (eds), *Multilevel Modelling of Health Statistics*. Chichester: Wiley.

Yang, S., Pickles, A., & Taylor, C. 1999. Multilevel Latent Variable Model for Analysing Two-phase Survey data. *Pages* 402-408 *of*: *Proceedings of the 14th International Workshop on Statistical Modeling*.

Yau, K. K. W., & Kuk, A. Y. C. 2002. Robust estimation in generalized linear mixed models. *Journal of the Royal Statistical Society, Series B*, 54, 101-117.

Yellott, J. 1977. The relationship between Luce's choice axiom, Thurstone's theory of comparative judgement, and the double exponential distribution. *Journal of Mathematical Psychology*, 15, 109-144.

Yule, G. U. 1912. On the methods of measuring association between two attributes. *Journal of the Royal Statistical Society*, 75, 579-652.

Yung, Y.-F. 1997. Finite mixtures in confirmatory factor-analysis models. *Psychometrika*, 62, 297-330.

Zeger, S. L., & Karim, M. R. 1991. Generalized linear models with random effects: A Gibbs sampling approach. *Journal of the American Statistical Association*, 86, 79-86.

Zeger, S. L., & Liang, K.-Y. 1986. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, 42, 121-130.

Zeger, S. L., Liang, K.-Y., & Albert, P. S. 1988. Models for longitudinal data: A generalized estimating equation approach. *Biometrics*, 44, 1049-1060.

Zellner, A. 1970. Estimation of regression relationships containing unobservable variables. *International Economic Review*, 11, 441-454.

Zellner, A. 1971. *An Introduction to Bayesian Inference in Econometrics*. New York: Wiley.

Zhao, L. P., & Prentice, R. L. 1990. Correlated binary regression using a quadratic exponential model. *Biometrics*, 77, 642-648.

Zimowski, M. F., Muraki, E., Mislevy, R. J., & Bock, R. D. 1996. *BILOG-MG 3: Multiplegroup IRT analysis and test maintenance for binary items*. Chicago, IL: Scientific Software International.

Zucchini, W. 2000. An introduction to model selection. *Journal of Mathematical Psychology*, 44, 41-61.

# 万卷方法总书目

万卷方法是我国第一套系统介绍社会科学研究方法的大型丛书，来自中国社科院、北京大学等研究机构和高校的两百余名学者参与了丛书的写作和翻译工作。至今已出版图书近80个品种，其中绝大多数是2007年以来出版的新书。

2010年版

1 社会科学定量研究的变量类型、方法选择及范例解析
978-7-5624-5174-5

2 案例研究：设计与方法
978-7-5624-5732-9

3 问卷设计手册：市场研究、民意调查、社会调查、健康调查指南
978-7-5624-5597-4

4 广义潜变量模型——多层次、纵贯性以及结构方程模型
978-7-5624-5393-2

5 调查问卷的设计与评估
978-7-5624-5153-2

6 心理学论文写作——基于APA格式的指导
978-7-5624-5354-3

7 心理学质性资料的分析
978-7-5624-5363-5

8 问卷统计分析实务：SPSS操作与应用
978-7-5624-5088-7

9 如何做综述性研究
978-7-5624-5375-8

10 质性访谈方法
978-7-5624-5307-9

11 量表编制：理论与应用（校订新译本）
978-7-5624-5285-0

12 质性研究：反思与评论（第2卷）
978-7-5624-5143-3

13 实验设计原理：社会科学理论验证的一种路径
978-7-5624-51877

14 混合方法论：定性研究与定量研究的结合
978-7-5624-5110-5

15 社会统计学
978-7-5624-5253-9

16 校长办公室的那个人（质性研究个案阅读）
978-7-5624-4880-8

17 泰利的街角（质性研究个案阅读）
978-7-5624-4937-9

18 客厅即工厂（质性研究个案阅读）
978-7-5624-48860

19 标准化调查访问
978-7-5624-5062-7

20 解释互动论
978-7-5624-49362

21 如何撰写研究计划书
978-7-5624-5087-0

22 质性研究的理论视角：一种反身性的方法论
978-7-5624-4889-1

23 社会评估：过程、方法与技术
978-7-5624-4975-1

24 如何解读统计图表
978-7-5624-4906-5

25 结构方程模型：AMOS的操作与应用
978-7-5624-4947-8

26 公共管理定量分析：方法与技术（第2版）
978-7-5624-3640-9

27 量化研究与统计方法
978-7-5624-4821-1

28 心理学研究要义
978-7-5624-5098-6

29 调查研究方法（第3版）
978-7-5624-3289-0

30 分析社会情境：质性观察和分析方法
978-7-5624-4690-3

31 建构扎根理论：质性研究实践指南
978-7-5624-4747-4

32 AMOS与研究方法
978-7-5624-4806-8

33 参与观察法
978-7-5624-46163

34 文化研究：民族志方法与生活文化
978-7-5624-46989

35 质性研究方法：健康及相关专业研究指南
978-7-5624-4720-7

36 如何做质性研究

978-7-5624-4697-2

37 质性研究中的访谈:教育及社会科学研究者指南
978-7-5624-4679-8

38 案例研究方法的应用(第2版)
7-5624-3278-3

39 应用 Stata 做统计分析(第5版)
978-7-5624-4483-1

40 教育研究方法论探索
978-7-5624-4649-1

41 实用抽样方法
978-7-5624-4487-9

42 质性研究:反思与评论(第1卷)
978-7-5624-4462-6

43 社会科学研究的思维要素(第8版)
978-7-5624-4465-7

44 哲学史方法论十四讲
978-7-5624-4446-6

45 社会研究方法
978-7-5624-4456-5

46 质性资料的分析:方法与实践(第2版)
978-7-5624-4426-8

47 实用数据再分析法(第2版)
978-7-5624-4296-7

48 质性研究的伦理
978-7-5624-4304-9

49 叙事研究:阅读、倾听与理解
978-7-5624-4303-2

50 质化方法在教育研究中的应用(第2版)
978-7-5624-4349-0

51 爱上统计学(第2版)
978-7-5624-4196-0

52 复杂调查设计与分析的实用方法(第2版)
978-7-5624-4290-5

53 美国心理协会写作手册(APA 格式)(第5版)
978-7-5624-4130-4

54 研究设计与写作指导:定性、定量与混合研究的路径
978-7-5624-3644-7

55 做自然主义研究:方法指南
978-7-5624-4259-2

56 多层次模型分析导论(第2版)
978-7-5624-4060-4

57 评估:方法与技术(第7版)
978-7-5624-3994-3

58 焦点团体:应用研究实践指南(第3版)
978-7-5624-3990-5

59 质的研究的设计:一种互动的取向(第2版)
978-7-5624-3971-4

60 组织诊断:方法、模型和过程(第3版)
978-7-5624-3055-1

61 民族志:步步深入(第2版)
978-7-5624-3996-7

62 分组比较的统计分析(第2版)
978-7-5624-3942-4

63 抽样调查设计导论(第2版)
978-7-5624-3943-1

64 定性研究(第3卷):经验资料收集与分析的方法(2版)
978-7-5624-3944-8

65 定性研究(第4卷):解释、评估与描述(第2版)
978-7-5624-3948-6

66 定性研究(第1卷):方法论基础(第2版)
978-7-5624-3851-9

67 定性研究(第2卷):策略与艺术(第2版)
978-7-5624-3286-9

68 社会网络分析法(第2版)
978-7-5624-2147-4

69 公共政策内容分析方法:
978-7-5624-3850-2

70 复杂性科学的方法论研究
978-7-5624-3825-0

71 社会科学研究:方法评论
978-7-5624-36898

72 论教育科学:基于文化哲学的批判与建构
978-7-5624-3641-6

73 科学决策方法:从社会科学研究到政策分析
7-5624-3669-0

74 电话调查方法:抽样、筛选与监控(第2版)
7-5624-3441-7

75 研究设计与社会测量导引(第6版)
978-7-5624-3295-1

为了建设好“万卷方法”，更好地服务学界，现由重庆大学出版社和人大经济论坛做出决定，凡购买重庆大学出版社的万卷方法系列图书的读者，填写以下信息调查表(复印即可)，邮寄给我们(400030 重庆大学出版社 林佳木)，经过认证后，我们将会赠送人大经济论坛币100个(可免费下载丛书相关学习资料并与教师及学友进行交流)：

| 读者情况调查表 | |
|---|---|
| 姓　名 | |
| 单　位 | |
| 联系电话 | |
| E-mail | |
| 论坛 ID | |
| 使用书籍 | |
| 购买渠道 | |
| 对丛书建设的建议 | |
| 邮政地址(邮编) | |

# 人大经济论坛

## ——国内最大的经济、管理、金融、统计类在线教育网站

人大经济论坛(网址:http://www.pinggu.org)依托中国人民大学经济学院，于2003年成立，致力于推动经济学科的进步，传播优秀教育资源，目前已经发展成为国内最大的经济、管理、金融、统计类的在线教育和咨询网站，也是国内最活跃和最具影响力的经济类网站。

1. 拥有国内经济类教育网站最多的关注人数，注册用户以百万计，日均数十万经济相关人士访问本站。

2. 是国内最丰富的经管类教育资源共享数据库和发布平台。

3. 论坛给所有会员提供学术交流与讨论的平台，同时也有网络社交SNS的空间，经管百科提供了丰富专业的经管类在线词典，数据定制和数据处理分析服务是您做实证研究的好帮手，免费的经济金融数据库使您不再为数据发愁，更有完善的经管统计类培训和教学相关软件，只要您是学习、研究或从事经管类行业，人大经济论坛就能满足您的需要！